

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 10:04:34 ; Search time 23252 Seconds
(without alignments) 525.683 Million cell updates/sec

Title: US-09-696-664A-3

Perfect score: 420
Sequence: 1 gcgcgtgtgagaattccta.....tcattgtgtgtaactattt 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vl:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pla:*
35: em.htg.rtd:*
36: em.htg.man:*
37: em.htg.vic:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	408.6	97.3	140384	8	ZMA86563	X86563 Zea mays co
2	399.4	95.1	1990	8	CHNRBCL	X55823 Neuracine m
3	399.4	95.1	2010	8	CHNRBCL	X55827 Neuracine t
4	392.2	93.4	1990	8	STICRBL	X79800 S.italica l
5	389.8	92.8	183990	8	AF114171	AF114171 Sorghum b
6	386	91.9	1803	8	CHZM02	V00171 Zea mays ch
7	385	91.7	1989	8	CHOSRBL	X04789 Rice chloro
8	385	91.7	2062	8	RICCPBCL	D00207 Oryza sativ
9	385	91.7	125936	8	AC122148	AC122148 Oryza sat
10	385	91.7	134525	8	CHOSYX	X15901 Oryza sativ
11	385	91.7	134933	8	AC092750	AC092750 Oryza sat
12	385	91.7	147201	8	AP003280	AP003280 Oryza sat
13	383.4	91.3	105649	2	AP003986	AP003986 Oryza sat
14	383.4	91.3	139914	2	AP004236	AP004236 Oryza sat
15	383.4	91.3	167379	2	AP004236	AP004236 Oryza sat
16	380.2	90.5	139671	2	CMS07EXT	AL513004 Oryza sat
17	377	89.8	1355	8	RICCPBCL	IC24073 Oryza sativ
18	376.6	89.7	749	8	AT090549	AY090549 Deschamps
19	376.6	89.7	1623	8	ASTCPBCL	L15300 Avena sativ
20	370.2	88.1	1431	8	CHZMRBCL	Z11973 Z.mays chlo
21	368.6	87.8	1755	8	WHCPBCL	D00206 Trifolium ae
22	368.6	87.8	2174	8	CHHV02	X00630 Barley chlo
23	368.6	87.8	134545	8	AB042240	AB042240 Trifolium
24	357.4	85.1	1510	8	CERRBCL	L14632 Penicillium se
25	354.2	84.3	1430	8	PENCARBCL	L15300 Avena sativ
26	354.2	84.3	2145	8	CHCLARBCL	X59731 C.laevigatu
27	348.8	83.0	2126	8	CHPCORBCL	X69740 P.comosa ch
28	345.6	82.3	1839	8	CHPRBCL	X55830 Flaveria bl
29	345.6	82.3	1842	8	CHPRBCL	X55829 Flaveria pr
30	344	81.9	2088	8	MTAPBCL	X83966 A.punctatu
31	344	81.9	2283	8	CHHRBCL	X15886 Cotton chlo
32	342.4	81.5	163935	8	MTSARBCL	X83996 S.australe
33	342.4	81.5	2087	8	OEL271079	AU271079 Oenothera
34	341.4	81.3	2193	8	IPRBCL	X60663 I.purpurea
35	340.8	81.1	1661	8	IECPBCL	X68583 Ilex vomito
36	340.8	81.1	2052	8	TOBPPBCL	M16867 Tobacco (N.
37	340.8	81.1	2524	8	TOBPPBCL	M16896 Tobacco (N.
38	340.8	81.1	4260	6	AX137516	J01450 Nicotiana t
39	340.8	81.1	4260	6	AX137516	AX137516 Sequence
40	340.8	81.1	4260	6	ES1181	ES1181 Method for
41	340.8	81.1	155939	8	CHNTXX	E51189 Process for
42	340.2	81.0	1344	8	TRU31381	Z00044 Nicotiana t
43	340.2	81.0	2074	8	CHNRBCL	U31381 Tristachya
44	339.8	80.9	2074	8	CHNRBCL	X69747 N.micronatu
45	339.2	80.8	2172	8	CHNRBCL	X69740 L.africana

ALIGNMENTS

RESULT 1
ZMA86563
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ZMA86563
140384 bp DNA
circular PLN 12-APR-2001
Zea mays complete chloroplast genome.
X86563.2 GI:11990232
16S rRNA, 23S rRNA, 4.5S rRNA, 5S rRNA, ATPase, atpE gene, atpI
gene, cema gene, clp gene, cytochrome b/f complex, cytochrome b6/
cytochrome f, infA gene, initiation factor 1, junction LSC-IR, matK
gene, NADH dehydrogenase, ndhB gene, ndhC gene, ndhD gene, ndhE
gene, ndhJ gene, ndhK gene, petA gene, petB gene, petD gene, petE
gene, psaA gene, psaB gene, psbA gene, psbB gene, psbD gene, psbJ
gene, psbK gene, psbN gene, PSI p700 apoprotein A1, PSI p700
apoprotein A2, PSII component, PSII cytochrome b559, PSII D2

PUBMED 1732000
REFERENCE 18 (bases 1 to 140384)
AUTHORS Rodermeil, S.R.
TITLE Nucleotide sequence of a maize chloroplast DNA fragment containing an inversion breakpoint, trng (GCC), trng (DCC), trnM, and a trng pseudogene

JOURNAL Nucleic Acids Res. 20 (21), 5844 (1992)
MEDLINE 93087206
PUBMED 1454547
REFERENCE 19 (bases 1 to 140384)
AUTHORS Maier, R.M., Neckermann, K., Hoch, B., Akhmedov, N.B. and Kossel, H.
TITLE Identification of editing positions in the ndh transcript from maize chloroplasts reveals sequence similarities between editing sites of chloroplasts and plant mitochondria

JOURNAL Nucleic Acids Res. 20 (23), 6189-6194 (1992)
MEDLINE 93117088
PUBMED 1282235
REFERENCE 20 (bases 1 to 140384)
AUTHORS Meglöhner, W. and Subramanian, A.R.
TITLE Nucleotide sequence of maize chloroplast rpl32: completing the apparent set of plastid ribosomal protein genes and their tentative operon organization

JOURNAL Plant Mol. Biol. 21 (3), 543-548 (1993)
MEDLINE 93184210
PUBMED 8443346
REFERENCE 21 (bases 1 to 140384)
AUTHORS Maier, R.M., Neckermann, K., Igloi, G.L. and Kossel, H.
TITLE Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing

JOURNAL J. Mol. Biol. 251 (5), 614-628 (1995)
MEDLINE 95395841
PUBMED 7666415
REFERENCE 22 (bases 1 to 140384)
AUTHORS Neckermann, K.
TITLE Direct Submission

JOURNAL Submitted (25-APR-1995) K. Neckermann, Institut f. Biologie III, Schenckstr. 1, D-79104 Freiburg i.Br., FRG
REMARK revised by (22)
REFERENCE 23 (bases 1 to 140384)
AUTHORS Neckermann, K.

Query Match 97.3%; Score 408.6; DB 8; Length 140384;
Best Local Similarity 99.0%; Pred. No. 1.4e-104;
Matches 411; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

6 TTGTGAGATCTTAAATCATGAGCTTGTAGGAGGAGGACTATGTCACCAACAAGAAAC 65
|||||
Db 56834 TTGTGAGATCTTAAATCATGAGCTTGTAGGAGGAGGACTATGTCACCAACAAGAAAC 56893

66 TAAACAGATGTGGATTAAAGTGTGTTAAGATATATAATGACTTACTACACCC 125
|||||
Db 56894 TAAACAGATGTGGATTAAAGTGTGTTAAGATATATAATGACTTACTACACCC 56953

126 GGAGTAGCAAAACAGATAGTATCTTGGCAGACTTCCGAGTACTCTCCAGCTCGG 185
|||||
Db 56954 GGAGTAGCAAAACAGATAGTATCTTGGCAGACTTCCGAGTACTCTCCAGCTCGG 57013

186 GGTTCGGCTGGAAGAAGAGAGCTGCGAGTCCGGAATCTTACTGTAACATGAGAC 245
|||||
Db 57014 GGTTCGGCTGGAAGAAGAGAGCTGCGAGTCCGGAATCTTACTGTAACATGAGAC 57073

246 AACTGTTGGACTGATGAGCTTACAGTCTGATGCTTACAAAGAGCATGATACAT 305
|||||
Db 57074 AACTGTTGGACTGATGAGCTTACAGTCTGATGCTTACAAAGAGCATGATACAT 57133

306 CGAGCCCGTCCCTGGGAGCAAGATCAATATCTGTTATGAGCTTATTCATTAAGACT 365
|||||
Db 57134 CGAGCCCGTCCCTGGGAGCAAGATCAATATCTGTTATGAGCTTATTCATTAAGACT 57193

366 ATTTGAAGGGTCTGTACTACATGTTCTTCACTTGGGTAACGATTT 420
|||||
Db 57194 ATTTGAAGGGTCTGTACTACATGTTCTTCACTTGGGTAACGATTT 57248

RESULT 2
CHNRBCL 1990 bp DNA linear PLN 29-MAR-2001
LOCUS subunit (nc 4.1.1.39).
DEFINITION Neutachne munit chloroplast rbcL gene for the rubisco large subunit (nc 4.1.1.39).
ACCESSION X55828 M33797 X15921
VERSION X55828.1 GI:11750
KEYWORDS rbcL gene; ribulose biphosphate carboxylase; rubisco large subunit.

SOURCE
ORGANISM Neutachne munit.
Plastid Neutachne munit
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Paniceae; Neutachne.

REFERENCE 1 (bases 1 to 1990)
AUTHORS Hudson, G.S., Mahon, J.D., Anderson, P.A., Gibbs, M.J., Badger, M.R., Andrews, T.J. and Whitfield, P.R.
TITLE Comparisons of rbcL genes for the large subunit of ribulose-bisphosphate carboxylase from closely related C3 and C4 plant species

JOURNAL U. Biol. Chem. 265 (2), 808-814 (1990)
MEDLINE 90110139
PUBMED 2295620
REFERENCE 2 (bases 1 to 1990)
AUTHORS Hudson, G.S.
TITLE Direct Submission

JOURNAL Submitted (20-JUL-1989) Hudson G.S., Research School of Biological Sciences, PO Box 475, Australian National University, Canberra City ACT 2601, Australia

FEATURES
source
1..1990
/organism="Neutachne munit"
/organelle="plastid"
/db_xref="taxon:4524"
87..92
/note="35 box"
111..116
/note="10 box"
121..>121
/note="put. rbcL transcript"
418..1854
/note="rubisco large subunit"
/codon_start=1
/transl_table=11
/protein_id="CAA39354.1"
/db_xref="GI:11751"
/db_xref="SWISS-PROT:P19163"
/translation="MSPQETKASVGFKAQKDYKLYTTPYEYTKDIDLAERVTP
OPGVPEEAGAAVALESSTGTATTTWDTLSIDRYKGRCHIEPVEEDQICVYA
YPLDEEBSVTNMTSTIVNGEFGALRLIEDLRIPYKTFQSPHIGYERD
KLVKGRPLIGCTIKPKLGLSAKNGRACVLELRGLDPTKDEWNSQPEWRMDRF
VCAEAIVYKAQAEIGIKNGHYLNATAGCTCEEMKRAAFREIGVPIWHDYTGSEFA
NLSMTCKDNGLLHIRAMAVIDRQNHGIRHVRVLAALRMGSGDHVHSGTVYVK
LGEERITLGFVDLNDLIERDNRKGVFTQDWVSMPEVITVASGGIHWVMPLTE
IFEDDVLVLEFGGTLGHPWGNPAGVANNVLAECVQANREGRDLARENELLKAAK
WSPETLAACEVWMAIKFEFEPEVDYDKV"
1941..1982
repeat_unit
BASE COUNT 613 a 340 c 445 g 592 t
ORIGIN

Query Match 95.1%; Score 399.4; DB 8; Length 1990;
Best Local Similarity 97.4%; Pred. No. 4.6e-102;
Matches 406; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

4 CGTTGTGAGATCTTAAATCATGAGCTTGTAGGAGGAGGACTTATGTCACCAACAAGAA 63
|||||
Db 376 CGTTGTGAGATCTTAAATCATGAGCTTGTAGGAGGAGGACTTATGTCACCAACAAGAA 435

64 ACTAAGCAAGTGTGGATTAAAGCTGCTTAAGGATATATAATGACTTACTACAC 123
|||||

Db 436 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGATTATTAATGACTTACTACACC 495
 QY 124 CCGAGTACGAAGCAAGCAAGTACTGATATCTGGCAGCAGTCCGAGTAACCTCAGCTC 183
 Db 496 CCGAGTACGAAGCAAGCAAGTACTGATATCTGGCAGCAGTCCGAGTAACCTCAGCTC 555
 QY 184 GGGGTCCGCTGGAAGAGCAGAGCTGAGTACTGCGGATCTTCTACTGATACAG 243
 Db 556 GGGGTCCGCTGGAAGAGCAGAGCTGAGTACTGCGGATCTTCTACTGATACAG 615
 QY 244 ACAACTGTTGGAGTACGATGACATACAGTCTGATGCTTCAAGAGCAGTGTATAC 303
 Db 616 ACAACTGTTGGAGTACGATGACATACAGTCTGATGCTTCAAGAGCAGTGTATAC 675
 QY 304 ATCAGCCCGTCTGCGGAGCAGATCAATATCTGATGATGAGCTTATCAGTACAG 363
 Db 676 ATCAGCCCGTCTGCGGAGCAGATCAATATCTGATGATGAGCTTATCAGTACAG 735
 QY 364 CTATTTGAAGAGGCTCTGCTACTACATGTTTCTTCTGATGAGTATTT 420
 Db 736 CTATTTGAAGAGGCTCTGCTACTACATGTTTCTTCTGATGAGTATTT 792
 LT 3
 CHNTRBCL
 LOCUS 2010 bp DNA linear PLN 29-MAR-2001
 DEFINITION Neurachne tenuifolia chloroplast rbcL gene for the rubisco large subunit (EC 4.1.1.39).
 ACCESSION X5827 M33796 X15920
 VERSION X5827.1 GI:11797
 KEYWORDS rbcL gene; ribulose biphosphate carboxylase; rubisco large subunit.
 SOURCE Neurachne tenuifolia.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC 1 (bases 1 to 2010)
 REFERENCE 1 Hudson, G.S., Mahon, J.D., Anderson, P.A., Gibbs, M.J., Badger, M.R., Andrews, T.J. and Whitfield, P.R. Comparisons of rbcL genes for the large subunit of ribulose-bisphosphate carboxylase from closely related C3 and C4 plant species
 JOURNAL J. Biol. Chem. 265 (2), 808-814 (1990)
 MEDLINE 90110139
 PUBMED 2295620
 REFERENCE 2 (bases 1 to 2010)
 AUTHORS Hudson, G.S.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-1989) Hudson G.S., Research School of Biological Sciences, PO Box 475, Australian National University, Canberra City ACT 2601, Australia
 FEATURES
 source
 1. 2010
 /organism="Neurachne tenuifolia"
 /organelle="plastid"
 /db_xref="taxon:4525"
 87..92
 /note="35 box"
 111..116
 /note="10 box"
 121..121
 /note="put. rbcL transcript"
 438..1874
 /note="rubisco large subunit"
 CDS
 precursor_RNA
 121..121
 /note="put. rbcL transcript"
 438..1874
 /note="rubisco large subunit"
 /translation="MSPTETKASGVKAGKDYKLYYYPEYETKDTLLAARVVP
 QPGVPEEAGAAVAASSTGTWTDGLSLDRYKGRCHIEPVGEEEOYICVYA
 YPLDFEBSVTNMTSIVGVFGFKALRALRLDLRIPIPTYSKTFQGPFGIGVEND

KLNKYPRLGCTIKPKILGSAKNYGRACYBOLRGIDFTKDNENVSOPMRRDRF
 VECARAIYKAQETGEIGKHYLNATAGCEEMRRAIRARLGIPIYVHYLITGFTTA
 NTSIAMGCRDNLHITRAHNAVVDROKNHMRFRVLAIRNSGGDHVHAGTVAK
 LEGEREITLGEVDLLRDP IERDRSGVFPQDVMSPGVIPVSGGIHVHMPALTE
 IFGDDSVLPFGGGTGHFWGNAPGAANRAVLAECVORANRGRDLARSGEIIKAACK
 MSPELAAACEVWKAIRKFEFAPVDYDKV"
 1961..2002
 repeat_unit
 /rpt_type=INVERTED
 BASE COUNT 614 a 345 c 453 g 598 t
 ORIGIN
 Query Match 95.1%; Score 399.4; DB 8; Length 2010;
 Best Local Similarity 97.4%; Pred. No. 4.6e-102;
 Matches 406; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 4 CGTTGGAGAAATCTTAATTCATGAGTGTAGGAGGAGCTTATGTCACCAACAGAA 63
 Db 396 CGTTGGAGAAATCTTAATTCATGAGTGTAGGAGGAGCTTATGTCACCAACAGAA 455
 QY 64 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGATTATTAATGACTTACTACACC 123
 Db 456 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGATTATTAATGACTTACTACACC 515
 QY 124 CCGAGTACGAAGCAAGCAAGTACTGATATCTTGGAGCAATTCGAGTACTCAGCTC 183
 Db 516 CCGAGTACGAAGCAAGCAAGTACTGATATCTTGGAGCAATTCGAGTACTCAGCTC 575
 QY 184 GGGGTCCGCTGGAAGAGCAGAGCTGAGTACTGCGGATCTTCTACTGATACAG 243
 Db 576 GGGGTCCGCTGGAAGAGCAGAGCTGAGTACTGCGGATCTTCTACTGATACAG 635
 QY 244 ACAACTGTTGGAGTACGATGACATACAGTCTGATGCTTCAAGAGCAGTGTATAC 303
 Db 636 ACAACTGTTGGAGTACGATGACATACAGTCTGATGCTTCAAGAGCAGTGTATAC 695
 QY 304 ATCAGCCCGTCTGCGGAGCAGATCAATATCTGATGATGAGCTTATCAGTACAG 363
 Db 696 ATCAGCCCGTCTGCGGAGCAGATCAATATCTGATGATGAGCTTATCAGTACAG 755
 QY 364 CTATTTGAAGAGGCTCTGCTACTACATGTTTCTTCTGATGAGTATTT 420
 Db 756 CTATTTGAAGAGGCTCTGCTACTACATGTTTCTTCTGATGAGTATTT 812
 RESULT 4
 SILENBCL
 LOCUS S.italica L. chloroplast rbcL gene. 1990 bp DNA linear PLN 29-OCT-1996
 DEFINITION X79900
 ACCESSION X79900.1 GI:1654091
 KEYWORDS large subunit; rbcL gene; ribulose 1,5-bisphosphate.
 SOURCE Setaria italica.
 ORGANISM Chloroplast Setaria italica
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC 1 (bases 1 to 1990)
 REFERENCE 1 Zhao, Y.S., Qiao, X.Y., Wu, N.H. and Wu, X.Y. Nucleotide sequence of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene from millet (Setaria italica)
 JOURNAL Acta Bot. Sin. 38, 719-724 (1996)
 REFERENCE 2 (bases 1 to 1990)
 AUTHORS Zhao, Y.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-1994) Y. Zhao, Institute of Developmental Biology, Academia Sinica, PO Box 2707, Beijing 100080, PRC
 FEATURES
 source
 1. 1990
 /organism="Setaria italica"
 /organelle="plastid:chloroplast"
 /db_xref="taxon:4555"
 /clone="psil"

repeat_region
/rpl_family="HCSR-1"
/rpl_type=dispersed
complement(22849..22086)
/rpl_family="HCSR-1"
/rpl_type=dispersed
complement(23211..22402)
/rpl_family="HCSR-1"
/rpl_type=dispersed
complement(join(<26698..27270,27261..27761,27976..28284,29824..31036,31131..31286,31385..31566,32230..32756,32998..33453,33728..34017,34138..34292,34882..35045))
/product="hypothetical protein"
complement(join(26698..27270,27361..27761,27976..28284,29824..31036,31131..31286,31385..31566,32230..32756,32998..33453,33728..34017,34138..34292,34882..35045))
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD27577.1"
/db_xref="GI:4680214"
/translation="MDGQQAARMAAQGVGVADVYAALRLQKFLAANDRRRLYD
DEPLRLAIRKACMLPLDLNRHTKAAVDSPLVPLDCEIMCHNLNVHLITMT
LCNSQLSVLIIGALFQVOYLIDCKKYYGRILANNVSEIQRKSLIQSK IMKELV
BEPLFLETKTSDYMDVNDVPGVAEDIIYDLSAVKROSPFYQGPPTKXGFLDEA
LARYAALFYLIKMNQEDJOREPVTQDIDLMTHTLOHPYQNDVATKLGVLEID
TDARSGKRLIDGFEETTEQFPENTGQVNRWKCMTGRKSPVSTQISTENG
TGSDCRKAQKMLADITAVELIYQIYDINNPNKENVYRFTKNQPDITSDGG
KLIDSTVGKNAAGVCLQCEPTEGLIVVMQDVSKKEPVGKVSFLQDILGDSKLS
EKEMKELAHGCHATSPVPSLRVAASATVSPSEFQVSNMMEPFSLSCILPHSILK
ONMSWTFEYVVDGCEGLIRLQIREOKANCGMAAVRELGVYLSPKKQFOIAERKNW
TLKDSMLITGTGDSLLDLKGDNOLIKIYQGRLEERKCSHSEDSAVATVAKCA
AHPKRAVLLDITRESOLIVNDEMFLIPYTTSVLEMDGDGSKLIGAVYQKGA
VPGSTSWAVETQAVAGASVATPAQCGTCTAFSGKVYAGFAEHAASGASRTYV
SEENHTSASGCGGCGSSCGPVAEDSRADNKSQGGCGCGGCGGGCGGCTMLK
ASTMASBEQSSRGSGCGGCGGGCGGCTMLKASTMASBEQSKSGCGCGGCGGCTMLK
GGCGGMAIEESKTLANAKSGGCGGCGGCGGCGGCTMLKASTMASVQSGSGCGG
GGCGGCGGCGGGGALITGSKTANAKSGGCGGCGGCGGCGGCTMLKSSSTADQHS
RSGGCGGCGGCGGGCGGSMVAEGSTGRNKSQGGCGGCGGCGGCGGCTTNNSTK
AGBGRKSHVSSGCGGCVLAVYRLVETSRGTSIANSGBERSADATPAAREDDSG
QTRRRAGGSGGSLAKRLIARVSTILKILARDGRESSEDRIVARSTLYLSTSPWT
PASSVGTPLFSGNVSAQLODVESQSPDVTGVRKAKGCTSHROKQPIEDKYIC
SAMLVNDSPIINGANOGSTFPWGRIRAFEPHKKRKAQVACRTSILMHMLITORDY
NKEFCGYSYKIERRNASGATIDDMINALEYITGADEENKFEPTLHCTLKIKEDKMS
LRIELQCNKKKQTKTKESTPNVSNNDDETPVAADSPERRRQGOVROKORDD
ASIANLEMLKKFARELERKAKAELSLEIFRASLEEKALLETKKRLSNEKKITDA
BLMKBEKIMLADKTSIDDEROLQWLMKNKILARHGN"
36421..36587
/rpl_family="HCSR-4"
/rpl_type=dispersed
complement(join(<8130..38275,38407..38612,38705..38895,39620..39776,39932..40049,40120..40673,40764..40864,41122..41697,41780..42115,42344..42727,42847..43083))
/product="hypothetical protein"
complement(join(38130..38275,38407..38612,38705..38895,39620..39776,39932..40049,40120..40673,40764..40864,41122..41697,41780..42115,42344..42727,42847..43083))
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD27562.1"
/db_xref="GI:4680199"
/translation="MKTSAICLARSVAEFLPIVGLFNLLILLOKSKANROVKOOL
LPKSNFYKSTISKAPPCCFHRLNDQYGENYPLPPGSSAATISTANASPTATPAAR
RPASQASPAVATPLRRHGKDWLGLFRRIRCLLDLGRGPTGRGCTVAAGTSSSD
RGCISTRRLRLILRRGKSYSDAAASLPPDDRILPKDSRSLGRIELTFLGSSG
KDAWMSDFENLFDACKKEIDAGNPMGLIYDQKEMWNLKRGNDSPGKRTKOLK
KLIDIKREYVEMEKNPATGLGDMAENKIVTQKEMWNLKRGNDSPGKRTKOLK
FRKRGKPTLEMHYIIPKKAHYDSTATCPDSDISDDSDVAEYKPTREKTYNQ
GKRKRGSTIAEERKSPFRVATKATCLIEAMERISTSEVSAAPLNTIPISA
EAVKAVKECGEESTALMYTAMSLANDEREFTSLETNNGDLORHEKERNKK
EGSMDIHLRAVAKGOKGILLALEYLSCHAAIYVGGNSYETKRMPSLEFVNLADLIY
SNYGTSDHLRAVAKGOKGILLALEYLSCHAAIYVGGNSYETKRMPSLEFVNLADLIY
AKNDITPDPDPESTHEHVRBNRPWPKFACALIDSSHVYIYVLENYHNHRCGY
TSQVAVIACDPDAFTVAYVGPSSAHSIILHALNALSPFVAERKAYLIVDSGTA
NRNGIAPDKSTYHIEFRRNRSEPPGKNMRKREDFRCADDEYKGRARRTOTHE

DGEDEGSDGDMNIIIRRIYDALTKETARRARRADESLTRGGGSRSPQAVRISGAL
 RRPPLVSGGRRRRRRAAGGAVVQSHQVPRCTAHNIQDEPRLVQGGIQOVAHVQVRA
 AEREGRRRRRRRVRGGRGGRGLVRLDRRRGRHRRRLPEELRRGSGADAGARRGRAR
 RRPRLRLPAVPRRLRLAARMLRHRTAM"
 join(<43647. .43308.43394. .44222.44944. .>45484)
 /product="hypotheetical protein"
 join(43647. .43908/43994. .44222.44944. .45484)
 /codon_start=1
 /product="hypotheetical protein"
 /protein_id="AAD27563.1"
 /db_xref="GI:4680200"
 /translation="MSGRIODSDSLIMLSVRLIPRYKOYLSFLFGMSHGIRCOHPSID
 LVRLPELSTREOGRIIDVLEEDHCRLISTGDTVTAQEAQRADAWGLPPR
 LSLDNLSEFFAAVALLPESRAALADTARRSCSLAAGSSTAKSSSTLKRPSS
 IDRASETTPTVYIEBLTGMEKADVDQDETRSLRLRQPARAAVNFKIDETVPSR
 DNARLSDPNTQINIESPSPEVFLVQIAPRPPLPSPVCSRLCRGAAVNAALCVMRG
 KRCRRRRDNTGTROMLLEMPADYVTHVTVIYDDDSIIHALLVLLACGLTALC
 LTVLRHGLT"
 complement(45613. .45755)
 /rpt_family="HCSR-4"
 /rpt_type=dispersed
 complement(join(446323. .46753,47300. .47458,48167. .>48248)))
 /product="hypotheetical protein"
 complement(join(46323. .46753,47300. .47458,48167. .48248))
 /note="similar to Oryza sativa EST clone C3025_1A"
 /codon_start=1
 /product="hypotheetical protein"
 /protein_id="AAD27561.1"
 /db_xref="GI:4680198"

Query Match	92.8%;	Score 389.8;	DB 8;	Length 18390;
Best Local Similarity	95.9%;	Pred. No. 2.9e-99;		
Matches 400;	Conservative 0;	Mismatches 17;	Indels 0;	Gaps 0;

QY	4	CGGTGAGAAATTCCTTAATTCATGAGTGTGAGGAGGAGCACTATGTCACCCAAACAGAA	63
Db	101567	CGTGTGAGAAATTCCTTAATTCATGAGTGTGAGGAGGAGCACTATGTCACCCAAACAGAA	101508
QY	64	ACTAAACCAAGTGTGATTTAAAGCTGGCTGTAAGCATTTAAATGACCTACTACAC	123
Db	101507	ACTAAACCAAGTGTGATTTAAAGCTGGCTGTAAGCATTTAAATGACCTACTACAC	101448
QY	124	CCGAGTAGCAAAACCAAGATACTGATTCCTGGCAGCATTCGAGTAGTACTCTCAGCTC	183
Db	101447	CCGAGTAGCAAAACCAAGATACTGATTCCTGGCAGCATTCGAGTAGTACTCTCAGCTC	101388
QY	184	GGGGTTCGGCTCAAAAGCAGAGCTGAGTGCAGCTGGCGAAATCTTCACTGGATCATG	243
Db	101387	GGGGTTCGGCTCAAAAGCAGAGCTGAGTGCAGCTGGCGAAATCTTCACTGGATCATG	101328
QY	244	ACAACCTGTTGGAGCTGATGAGCTTACCAAGTCTTGATCTCTTCAAAAGACAGATGCTATCAC	303
Db	101327	ACAACCTGTTGGAGCTGATGAGCTTACCAAGTCTTGATCTCTTCAAAAGATGATGCTATCAC	101268
QY	304	ATCGAGCCGCTTCCTGGGAGCCCAAGTCAATATATCTGTATGTAGCTTATCCATTAGAC	363
Db	101267	ATCGAGCCGCTTCCTGGGAGCCCAAGTCAATATATCTGTATGTAGCTTATCCATTAGAC	101208
QY	364	CTATTGTAAGAGGCTCTGTACTAATCAATGTTTACTTTCATTGCGGTAACGTAATTT	420
Db	101207	CTATTGTAAGAGGCTCTGTACTAATCAATGTTTACTTTCATTGCGGTAACGTAATTT	101151

RESULT 6

CH2M02

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CH2M02

1803 bp

DNA

linear

PLN 29-MAR-2001

Ze mays chloroplast gene for the large subunit of RUBP (ribulose biphosphate carboxylase).

V00171 J01423

V00171.1 GI:12394

ribulose biphosphate carboxylase.

Ze mays.

Chloroplast Zea mays

REFERENCE

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC clade: Panicoideae: Andropogoneae: Zea.

AUTHORS

McIntosh, L., Poulsen, C. and Bogorad, L.

TITLE

Chloroplast gene sequence for the large subunit of ribulose biphosphatocarbonylase of maize

JOURNAL

Nature 288, 556-560 (1980)

COMMENT

On Jul 24, 2002 this sequence version replaced gi:342623.

FEATURES

location/Qualifiers

CDS

1..1803
/organism="Zea mays"
/organelle="plastid:chloroplast"
/db_xref="taxon:4577"
228..1655
/codon_start=1
/product="ribulose biphosphatase carboxylase"
/protein_id="CAA23474.1"
/db_xref="GI:12395"
/translation="MSPQETKASVGFKAGVDKLYTPPEYETKDTLLAARVTP
OLGYPPEAGAAVAEASAGTWTWTGTLSDRYKGCYHIEPVPDQYICV
AYPLDEEGSVNMTSTIYGNVEFKALALDELRIIPYKSTFOGPRMOYER
DKLNKGRPLIGCTIKRKLSSKNYGRACEEIRGDLDTKDDENNSPPYKMRDR
FVFCALYKQAEETIKGHYLNATGCTDEMTKGVAFARQLGVPLVMHDYLTGFT
ANTLLSHYCRDNGLLIHRAMAVIDRKNHGMFRYLAALRMSGDHHSYVVK
LEGREITLGFVDLIRDDPTEKDRSGIEFTQWVSMPEVYIPVAGGILHVMHMPALTE
ILGDSVLOFGGTLGHPMGNAGAAANRVALACVQARNREGDLAREVQIIKACRW
SAELAAACEIMKEIKFEPEPKMDTI"

BASE COUNT 539 a 327 c 410 g 527 t
ORIGIN

Query Match 91.9%; Score 386; DB 8; Length 1803;
Best Local Similarity 96.9%; Pred. No. 2.8e-98;
Matches 405; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

OY 6 TTGTGAGAAATTTTAAATTCATGAGTTGAGGAGGAGCTTATGTCACCAACAGAAC 65
DB 188 TTGTGAGAAATTTTAAATTCATGAGTTGAGGAGGAGCTTATGTCACCAACAGAAC 247
OY 66 TAAACAGAGTGTGATTTAAAGCTGTGTTAAGGATTAATTAATGACTTACTACACCC 125
DB 248 TAAACAGAGTGTGATTTAAAGCTGTGTTAAGGATTAATTAATGACTTACTACACCC 307
OY 126 GGAGTACGAAACCAAGATGATGATATCTTGGACACATTCGAGTACTCTCCAGCTCGG 185
DB 308 GGAGTACGAAACCAAGATGATGATATCTTGGACACATTCGAGTACTCTCCAGCTCGG 367
OY 186 GGTTCGGCTGAGAGAGAGAGCTGACAGTACGCTGCGAATCTT--CTACTGGTACATG 242
DB 368 GGTTCGGCTGAGAGAGAGAGCTGACAGTACGCTGCGAATCTTCTGCTGGTACATG 427
OY 243 GACAACTGTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 302
DB 428 GACAACTGTTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 487
OY 303 CATGAGAGCCGTTCTGAGGAGCCAGATCAATATATCTGTTATGAGTATTCATTAGA 362
DB 488 CATGAGAGCCGTTCTGAGGAGCCAGATCAATATATCTGTTATGAGTATTCATTAGA 547
OY 363 CCTATTGAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 420
DB 548 CCTATTGAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 605

RESULT 7

CHOSRBL

1989 bp DNA linear PLN 29-MAR-2001

DEFINITION Rice chloroplast DNA for ribulose-1,5-biphosphate carboxylase

ACCESSION

X04789
X04789.1 GI:119954

KEYWORDS

ribulose biphosphatase carboxylase.

SOURCE

Oryza sativa.

ORGANISM

Plastid Oryza sativa
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 (bases 1 to 1989)

AUTHORS

Moon, E., Kao, T.H. and Wu, R.

TITLE

Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes

JOURNAL

Nucleic Acids Res. 15 (2), 611-630 (1987)

MEDLINE

87146381

PUBMED

3023686

FEATURES

location/Qualifiers

1..1989
/organism="Oryza sativa"
/organelle="plastid"
/strain="labelle"
/db_xref="taxon:4530"
/clone="pct-1"
41..46
/note="pot. -35 region"
65..70
/note="pot. -10 region"
377..383
/note="pot. ribosome binding site"
388..1821
/note="ribulose-1,5-biphosphate carboxylase (AA 1 - 477)"
/codon_start=1
/translation="MSPQETKASVGFKAGVDKLYTPPEYETKDTLLAARVTP
OLGYPPEAGAAVAEASAGTWTWTGTLSDRYKGCYHIEPVPDQYICV
AYPLDEEGSVNMTSTIYGNVEFKALALDELRIIPYKSTFOGPRMOYER
DKLNKGRPLIGCTIKRKLSSKNYGRACEEIRGDLDTKDDENNSPPYKMRDR
FVFCALYKQAEETIKGHYLNATGCTDEMTKGVAFARQLGVPLVMHDYLTGFT
ANTLLSHYCRDNGLLIHRAMAVIDRKNHGMFRYLAALRMSGDHHSYVVK
LEGREITLGFVDLIRDDPTEKDRSGIEFTQWVSMPEVYIPVAGGILHVMHMPALTE
ILGDSVLOFGGTLGHPMGNAGAAANRVALACVQARNREGDLAREVQIIKACRW
SAELAAACEIMKEIKFEPEPKMDTI"

CDS

BASE COUNT 600 a 348 c 452 g 589 t
ORIGIN

Query Match 91.7%; Score 385; DB 8; Length 1989;
Best Local Similarity 95.2%; Pred. No. 5.4e-98;
Matches 397; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 4 CGTTGTGAGAAATTTTAAATTCATGAGTTGAGGAGGAGCTTATGTCACCAACAGAA 63
DB 346 CGTTGTGAGAAATTTTAAATTCATGAGTTGAGGAGGAGCTTATGTCACCAACAGAA 405
OY 64 ACTAAGCAAGTGTGATTTAAAGCTGTGTTAAGGATTAATTAATGACTTACTACAC 123
DB 406 ACTAAGCAAGTGTGATTTAAAGCTGTGTTAAGGATTAATTAATGACTTACTACAC 465
OY 124 CCGAGTACGAAACCAAGATGATGATATCTTGGACACATTCGAGTACTCTCCAGCTC 183
DB 466 CCGAGTACGAAACCAAGATGATGATATCTTGGACACATTCGAGTACTCTCCAGCTC 525
OY 184 GGGGTTCCGCTGAGAGAGAGAGCTGACAGTACGCTGCGAATCTTCTACTGGTACATG 243
DB 526 GGGGTTCCGCTGAGAGAGAGAGCTGACAGTACGCTGCGAATCTTCTACTGGTACATG 585
OY 244 ACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 303
DB 586 ACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 645
OY 304 ATGAGAGCCGTTCTGAGGAGCCAGATCAATATATCTGTTATGAGTATTCATTAGAC 363
DB 646 ATGAGAGCCGTTCTGAGGAGCCAGATCAATATATCTGTTATGAGTATTCATTAGAC 705

QY 364 CTAATTGAAGAGGCTTCTGTACTACATGTTTACTTCTTCTGCGGAGTATTT 420
 Db 706 CTAATTGAAGAGGCTTCTGTACTACATGTTTACTTCTTCTGCGGAGTATTT 762

RESULT 8
 LOCUS RICCPBCL 2062 bp DNA linear PLN 28-Apr-2001
 DEFINITION *Oryza sativa* chloroplast gene for ribulose 1,5-bisphosphate
 carboxylase large subunit, complete cds.
 ACCESSION D00207
 VERSION D00207.1 GI:344016
 KEYWORDS Rubisco LS; chloroplast; large subunit; rbcL; ribulose 1,
 5-bisphosphate carboxylase.
 SOURCE *Oryza sativa* chloroplast DNA.
 ORGANISM *Oryza sativa*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Erihartoideae; Oryzaceae; *Oryza*.
 REFERENCE 1 (bases 1 to 2062)
 AUTHORS Nishizawa, Y. and Hirai, A.
 TITLE Nucleotide sequence and expression of the gene for the large
 subunit of rice ribulose 1,5-bisphosphate carboxylase
 JOURNAL Jpn. J. Genet. 62, 389-395 (1987)
 COMMENT The initiation and termination sites of the Rubisco LS mRNA were
 also determined by SI mapping method.
 FEATURES
 source
 1. 2062
 /organism="Oryza sativa"
 /organelle="plastid:chloroplast"
 /db_xref="taxon:4530"
 /note="588 bp upstream of PstI site"
 74..79
 -35_signal
 -10_signal
 108..>1854
 411..415
 421..1854
 421..1854
 /gene="rbcL"
 /note="rbcL"
 /gene="Rubisco LS"
 /note="Rubisco LS"
 /transl_start=1
 /transl_table=1
 /product="ribulose 1,5-bisphosphate carboxylase large
 subunit"
 /protein_id="BA00147.1"
 /db_xref="GI:344017"
 /translation="MSPTETKASVGFKAQVNDKYLYTPEYETKNDIILAFRVT
 OPGVPEEAGAAVAASSSTGWTWTDLTSDIRKKGCHIEPYEGNDQYIAYVA
 YLDFDEBSVYMFISYGVNFGKRALEKRLDRIPTYSKTFQGPPIQVERD
 KLNKIGRPLIGCTIKPKLGLSKNKGRAECLEKGLDFTKIDENVNSQPFKMRDRF
 VFCAEIVYSOAEIETKIKGHYLNATGCEEMIKRAVEIRELGVIMYDITGCTFA
 NTSIAHCNDNGILLIHRAMHAVIDROKNHMFYVLAALRMGGGDIIHGTGTA
 LBEREMTGEPVLDLDDDEIERDRAAGIPEFODVSMPEVTIVASGGIHWMMPLATE
 IREGDSVLOFGGTLGHPWNGAAGAANVALAEACVQARNESRDILARENETIRACK
 MSPELAACEIKATIEFEPEPKDKDS"

BASE COUNT 626 a 358 c 468 g 610 t
 ORIGIN

Query Match 91.7%; Score 385; DB 8; Length 2062;
 Best Local Similarity 95.2%; Pred. No. 5.4e-98;
 Matches 397; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 CGTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGCTTATGTCACCAACAGAA 63
 Db 379 CGTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGGAGCTTATGTCACCAACAGAA 438
 QY 64 ACTRAAGCAAGTGTGATTTAAAGCTGCTTAAAGCATTAATTAATGACTTACTACACC 123
 Db 439 ACTRAAGCAAGTGTGATTTAAAGCTGCTTAAAGCATTAATTAATGACTTACTACACC 498
 QY 124 CCGAGTAGCAACCAAGATGATCTGATCTTGGAGCAATTCGATTAATCTCTACGCC 183

Db 499 CCGAGTAGCAACCAAGATGATCTGATCTTGGAGCAATTCGATTAATCTCTACGCC 558
 QY 184 GGGGTTCCGCTTAAGAAGAGAGAGTGCATGACGTGCGGAATCTTACTGTGATCAGG 243
 Db 559 GGGGTTCCGCTTAAGAAGAGAGAGTGCATGACGTGCGGAATCTTACTGTGATCAGG 618
 QY 244 ACAACGTGTTGACATGATGACCTTACAGCTGATGCTTTACAAAGAGAGATGATAC 303
 Db 619 ACAACGTGTTGACATGATGACCTTACAGCTGATGCTTTACAAAGAGAGATGATAC 678
 QY 304 ATGAGCCGCTTCTGCGGACCCAGATCAATATATCTGTTATGATGATTCATTAGAC 363
 Db 679 ATGAGCCGCTTCTGCGGAGAGATCAATATATGCTTATGATGATTCATTAGAC 738

QY 364 CTAATTGAAGAGGCTTCTGTACTACATGTTTACTTCTTCTGCGGAGTATTT 420
 Db 739 CTAATTGAAGAGGCTTCTGTACTACATGTTTACTTCTTCTGCGGAGTATTT 795

RESULT 9
 LOCUS AC122148
 DEFINITION *Oryza sativa* (japonica cultivar-group) chromosome 10 clone
 OSJNAB0075K12, complete sequence.
 ACCESSION AC122148
 VERSION AC122148.1 GI:21104881
 KEYWORDS HTG.
 SOURCE
 ORGANISM
Oryza sativa (japonica cultivar-group).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Erihartoideae; Oryzaceae; *Oryza*.
 REFERENCE 1 (bases 1 to 125936)
 AUTHORS Wang, R.A., Yu, Y., Yang, T.J., Nah, G., Soderlund, C., Chen, M.,
 Kim, H.-R., Rambo, T., Saski, C., Henry, D., Oates, R. and Simmons, J.
 TITLE Rice Genomic Sequence
 JOURNAL Unpublished
 2 (bases 1 to 125936)
 AUTHORS Wang, R.A., Yu, Y., Yang, T.J., Nah, G., Soderlund, C., Chen, M.,
 Kim, H.-R., Rambo, T., Saski, C., Henry, D., Oates, R. and Simmons, J.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Clemson University Genomics Institute,
 Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
 3 (bases 1 to 125936)
 REFERENCE 2
 AUTHORS Wang, R.A., Yu, Y., Yang, T.J., Nah, G., Soderlund, C., Chen, M.,
 Rambo, T., Saski, C., Henry, D., Oates, R., Simmons, J. and
 Thurmond, S.K.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUN-2002) Clemson University Genomics Institute,
 Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
 COMMENT The following sequence is an artificial clone that will be used for
 gap filling in the rice genome. There is 60 kb of sequence from
 the overlapping north clone (OSJNAB0075K12) and 60 kb of sequence
 from the overlapping south clone (OSJNAB0022D10) and the gap
 filling sequence begins at 60,001 and ends at 65,936 for a total
 size of 125,936 bases. The gap filling sequence is 5936 bases. The
 name (OSJNAB0075K12) was derived from the location of the north
 clone and the 'A' stands for Artificial.

FEATURES
 source
 1. 125936
 /organism="Oryza sativa (japonica cultivar-group)"
 /db_xref="taxon:39947"
 /chromosome="10"
 /clone="OSJNAB0075K12"
 /complement(<1..1590)
 /gene="OSJNAB0075K12.1"
 /note="3' Partial - Hypothetical protein ORF542 from
 chromosome 10 chloroplast insertion"
 /complement(<1..1590)
 /gene="OSJNAB0075K12.1"
 /note="3' Partial - Hypothetical protein ORF542 from
 chromosome 10 chloroplast insertion"
 repeat_region
 complement(2437..2462)

gene /note="low_complexity AT-rich"
complement(2781..3848)
/gene="OSJNAB0075K12.2"
/note="Putative ribosomal protein S16 from chromosome 10
chloroplast insertion"
complement(join(2781..2929,3809..3848))
/gene="OSJNAB0075K12.2"
/codon_start=1
/product="Putative ribosomal protein S16 from chromosome
10 chloroplast insertion"
/protein_id="AA048248.1"
/db_xref="GI:21327343"
/translation="MLKRLKRCGRQREYDPIKNOTCLNVPAILYFLEKGAQPRFV
SDLRKFEFEKERILS"
complement(4214..4238)
/note="Low_complexity AT-rich"
complement(4610..4644)
/note="low_complexity AT-rich"
5327..5512
/gene="OSJNAB0075K12.3"
/note="Putative PSII K protein from chromosome 10
chloroplast insertion"
5327..5512
/gene="OSJNAB0075K12.3"
/codon_start=1
/product="Putative PSII K protein from chromosome 10
chloroplast insertion"
/protein_id="AA048249.1"
/db_xref="GI:21327344"
/translation="MPNLSLTCICFNSVIYPTSPFEAKLPEAVFNPVDFMPV
VLFLAFIMQAAVSFR"
5901..6011
/gene="OSJNAB0075K12.4"
/note="Putative PSII I protein from chromosome 10
chloroplast insertion"
5901..6011
/gene="OSJNAB0075K12.4"
/codon_start=1
/product="Putative PSII I protein from chromosome 10
chloroplast insertion"
/protein_id="AA048250.1"
/db_xref="GI:21327345"
/translation="MLTKLFVTVVILFVSLFGLSNDPGKNGRDE"
6642..6875
/gene="OSJNAB0075K12.5"
/note="Hypothetical protein ORF100 from chromosome 10
chloroplast insertion"
6642..6875
/gene="OSJNAB0075K12.5"
/codon_start=1
/product="Hypothetical protein ORF100 from chromosome 10
chloroplast insertion"
/protein_id="AA048251.1"
/db_xref="GI:21327346"
/translation="MKLYKWLFLYLRPRKARISRYLLPHILVTRCSIGIRNRNP
FPIHKMFESQFFSISLFEIPLKNRLSMK"
8133..9554
/gene="OSJNAB0075K12.6"
/note="Putative PSII 43kDa protein from chromosome 10
chloroplast insertion"
8133..9554
/gene="OSJNAB0075K12.6"
/codon_start=1
/product="Putative PSII 43kDa protein from chromosome 10
chloroplast insertion"
/protein_id="AA048252.1"
/db_xref="GI:21327347"
/translation="MKILVSLRRFYHVEMLNGFTVLAGDQETTGFPAMAGNARLIN
LSGKLGAVAHAGLIVFAGAMLFEVAHFPEKPMTEGQILPLHATLGMVGPG
GEVLDPEFVSGVLHLSAVLFGGAYHALGPELTESFPFGYKADKRNKTTI
LGIHLIGLIGALFLVLRKALFGGIDTMAPGGGVRKTIINTLSPGVIFGLKSP
GGSGHIVSVDDLEDITIGKHWMLGFTICGCIWHILTRPAMAHRAVFGSEAYISYL
GALSVFGFIACCFWFMNNTAIVSSEFTGIPGPAISOQAFTPLVTRQRLGANVGSAGCP

gene TGGLKYIMRSPTGEVIFGGETMRFWDLAPWLEIRLPGNDLSRLKKDIOPMOERHS
AEYMTAPGLSINSVGVATLEINAVVSPRSMLATSHFVLGFFFPVGHLMHAGRARA
AAGFEKGLDRDLEPYLWTPPLN"
10161..10349
/gene="OSJNAB0075K12.7"
/note="Hypothetical protein ORF62 from chromosome 10
chloroplast insertion"
10161..10349
/gene="OSJNAB0075K12.7"
/codon_start=1
/product="Hypothetical protein ORF62 from chromosome 10
chloroplast insertion"
/protein_id="AA048253.1"
/db_xref="GI:21327348"
/translation="MTIAFOIAVFAVITVSVLVISVPEFVAFSPDGSNNKNVFSGT
SLMTGIVFAIINLSIS"
complement(12297..12566)
/gene="OSJNAB0075K12.8"
/note="Hypothetical protein ORF91 from chromosome 10
chloroplast insertion"
complement(12297..12566)
/gene="OSJNAB0075K12.8"
/codon_start=1
/product="Hypothetical protein ORF91 from chromosome 10
chloroplast insertion"
/protein_id="AA048254.1"
/db_xref="GI:21327349"
/translation="MYGYCKSNMHPNRRKGTQDYWEELLVMSGLYALFCVF
LVLFIPDSFOESNKLESGKEKKENENLSHDIONLYIK"
complement(12791..13003)
/gene="OSJNAB0075K12.9"
/note="Hypothetical protein ORF70 from chromosome 10
chloroplast insertion"
complement(12791..13003)
/gene="OSJNAB0075K12.9"
/codon_start=1
/product="Hypothetical protein ORF70 from chromosome 10
chloroplast insertion"
/protein_id="AA048255.1"
/db_xref="GI:21327350"
/translation="MKTFSIYVLPPLIGMGINFAPRIPYGVWLIIFSILVCGEGR
LKQPLTGIPMKNLVQNPALFKRIC"
14907..15011
/gene="OSJNAB0075K12.10"
/note="Putative PSII low MW protein from chromosome 10
chloroplast insertion"
14907..15011
/gene="OSJNAB0075K12.10"
/codon_start=1
/product="Putative PSII low MW protein from chromosome 10
chloroplast insertion"
/protein_id="AA048256.1"
/db_xref="GI:21327351"
/translation="MEVNILAFIAFALFIIIVPFAFLIIYVKVSOND"
complement(15774..15863)
/gene="OSJNAB0075K12.11"
/note="Hypothetical protein ORF29 from chromosome 10
chloroplast insertion"
complement(15774..15863)
/gene="OSJNAB0075K12.11"
/codon_start=1
/product="Hypothetical protein ORF29 from chromosome 10
chloroplast insertion"
/protein_id="AA048257.1"
/db_xref="GI:21327352"
/translation="MDIVSLAWALMVVTFESLSLVWGRSGL"
17155..17317
/note="Similar to Oryza sativa telomere-associated
sequence"
complement(20592..20717)
/note="Similar to Oryza sativa telomere-associated
sequence"
complement(27558..27579)

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

OSUNB0075K12, from chromosome 10, complete sequence
2 (bases 1 to 134933)
McCombie, W.R.
Direct Submission
Submitted (25-JUL-2001) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY 11724, USA
3 (bases 1 to 134933)
McCombie, W.R.
Direct Submission
Submitted (24-OCT-2001) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY 11724, USA
4 (bases 1 to 134933)
Palmer, L.E., Yu, M., de la Bastide, M., Spiegel, L., Nascimben, L., Zuluverna, T., Ballis, V., Bell, M., Preston, R., Kirchoff, K., Kull, R., Baker, J., Santos, L., Miller, B., Cummins, D.M., Katzenberger, F., Muller, S., Shah, R., King, L., Yang, C., Dike, S., O'Shaughnessy, A., Dechla, N. and McCombie, W.R.
Direct Submission
Submitted (04-APR-2002) Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Genomic sequence for *Oryza sativa* (japonica cultivar-group) cultivar Nipponbare clone OSUNB0075K12, from chromosome 10
On Oct 24, 2001 this sequence version replaced g1:15011683.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.
location/Qualifiers
1.134933
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="10"
/clone="OSUNB0075K12"
/clone_11b="EcoRI"
/note="japonica cultivar-group"
complement(20..226)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
Ty3-gypsy-like
124..2469
/gene="OSUNB0075K12.1"
/note="Putative retroelement"
join<124..958,1238..1497,1930..2469)
/gene="OSUNB0075K12.1"
/note="Similar to RETROVIRUS-RELATED POL POLYPROTEIN FROM TRANSPOSON 297 [CONTAINS]"
/codon_start=1
/product="Putative retroelement"
/protein_id="AA08567.1"
/db_xref="GI:19920135"
/translation="OHKLEAPSKCSFALQOLEYGHITTSKTVADPAKTYLWQMPVPSVLADKRGFLGLGYRRKRVKYYGLAPLIELEKKNIPFWTSYTOQAFETLQAMITTPVSLIDPSPKPVETIDCTGTGIDVLSQGHPIAYSAIGTANQKLSYREKFMATMAVDKRPYTLNRPVYIKTDHFLCHLDTHVDSMOKKAMKLIQLOVHYQYKGVNKSADALSRVGOEFACSSVSLVOPPIQIEIVSYADVQAOQLQOELAVISPNQGVWQDISIDIDGLPLSDSKNVLIVVDRFKEFAFEILKHPYNAOHAASFEDRAASIELEPTWSDVRALETSQFNKVLFAKKKSVASRPYKLAFTKEGPEILSKYGSVAVRLKLPVDSOVRPFVHVSQIKHVPVSHVPSSELPYQDIDSLPEPEILIDRLVKKGNALIVQVIRGLGVLPBCATREDEYSVLRAPPLATVGHVSHSGGSVYPPGVAPEDAGITQORREDELRKRRLPEVKAYLCNVCARA"

repeat_region
406..690
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(11303..1562)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"

repeat_region
complement(4245..4320)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(5779..5957)
/note="Simple-repeat (CCGGG)n"
6170..6328
/note="Simple-repeat (CCGGG)n"
complement(6495..6547)
/note="Low-complexity AT-rich"
7667..9036
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(9725..9807)
/note="Similar to unknown repetitive sequence"
complement(9908..10030)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(10048..10105)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(10336..10423)
/note="Similar to unknown repetitive sequence"
10468..10552
/note="Similar to unknown repetitive sequence"
complement(10568..10643)
/note="Similar to unknown repetitive sequence"
complement(10698..12424)
/gene="OSUNB0075K12.2"
/note="Hypothetical protein"
complement(join(10698..11211,11988..12424))
/gene="OSUNB0075K12.2"
/note="Similar to (AP002523) hypothetical protein similar to *Oryza sativa* chromosome 6, P0680A03.17 E-value=105 identities=210/732 (of subject)"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AA08568.1"
/db_xref="GI:19920136"
/translation="MDSRLSEPYDQDPVHNSAGLSPPPLITTTSPACLRRLVSPTRLCYAAVYGRRLRLHRLAYADAMCKRHGWTIMPLGVNIFGLPSPPTLSVAYVGVSHHRLKASSTPTDMPSPPTGVSAVYKSLCRNRIQILGTGKIGFKLISVAGDQSHHRLKLVAVADCFRLHGVNPPPLGVYFTDRLALVRLVPPAKLAVYAVGRLRLHRLPAASSADWCFCFLGWPIMPPLGVYDFDRPDRPPCVSACCTTILRCWASTSSPTGRLVRRQLVFPARLAYVDAN"

repeat_region
10704..10778
/note="Similar to unknown repetitive sequence"
10779..10867
/note="Similar to unknown repetitive sequence"
10868..10956
/note="Similar to unknown repetitive sequence"
10957..11047
/note="Similar to unknown repetitive sequence"
complement(11048..11133)
/note="Similar to unknown repetitive sequence"
complement(11932..11976)
/note="Similar to unknown repetitive sequence"
11978..12066
/note="Similar to unknown repetitive sequence"
12069..12156
/note="Similar to unknown repetitive sequence"
12157..12245
/note="Similar to unknown repetitive sequence"
12246..12334
/note="Similar to unknown repetitive sequence"
complement(13006..13517)
/note="Similar to putative retrotransposon, Ty3-gypsy-like"
complement(13638..13727)
/note="Similar to unknown transposon"
13781..13809
/note="Low-complexity AT-rich"
14700..14759
/note="Simple-repeat (CGGGG)n"

gene complement(15210..17749)
/gene="OSJNB0075K12.3"
/note="Hypothetical protein"
complement(join(15210..15407,17193..17329,17428..17749))
/gene="OSJNB0075K12.3"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AA008569.1"
/db_xref="GI:19920137"
/translation="MGNCQADAAAVVIOHPSSSSSSSSSGNGGGGGGGERAYGA
VSAAVVAAAPGHVAVVPPVATPAAPAAATASAPARRRLKLPDPTLVGVYR
LTFEEDVLOFARSKRNATATAGDDEDDGHRGGGGAAPAKYAAQSRQ
ENRSPSPDDEARPEPEPEPLVATMAAGRMELSHGGRAPLPSIAGSVLCF"

CDS

repeat_region
/note="Similar to putative MITE, MITE-adh, type B-like"
15903..16145
repeat_region
16234..16431
/note="Similar to putative MITE, Wanderer-like"
complement(16500..16527)
repeat_region
/note="Low complexity AT-rich"
17140..17161
/note="Simple-repeat (CATCT)n"
complement(17494..17545)
repeat_region
/note="Simple-repeat (CGG)n"
17625..17680
/note="We believe the assembly to be correct. The
assembly is covered by a single subclone on the top strand
with weak coverage by an additional subclone of less than
phred30 quality on the bottom strand. The area is being
confirmed by PCR off of the BAC DNA."
17670..17700
repeat_region
/note="Simple-repeat (CGA)n"
complement(18420..18534)
repeat_region
/note="Similar to putative MITE, MITE-adh-1-like"
complement(19588..23866)
gene
/gene="OSJNB0075K12.4"

Query Match 91.7% Score 385; DB 8; Length 134933;
Best local Similarity 95.2% Pred. No. 6.4e-98;
Matches 397; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 4 CGTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGCTTATGTCACCAACAGAA 63
Db 127196 CGTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGCTTATGTCACCAACAGAA 127255

64 ACTAAGCAAGCTGTGATTTAAAGCTGTGTAAGATTAATTAATGCTACTACAC 123
127256 ACTAAGCAAGCTGTGATTTAAAGCTGTGTAAGATTAATTAATGCTACTACAC 127315

OY 124 CCGAGTAGCAAGCAAGATGATATCTTGGCAGATTCGAGTAATCTCAGCTC 183
Db 127316 CCGAGTAGCAAGCAAGATGATATCTTGGCAGATTCGAGTAATCTCAGCTC 127375

OY 184 GGGGTTCCGCTGAGAGAGAGAGCTGCAGTAGCTGCGGATCTTCTCTGCTGATCAGG 243
Db 127376 GGGGTTCCGCTGAGAGAGAGAGCTGCAGTAGCTGCGGATCTTCTCTGCTGATCAGG 127435

OY 244 ACAAGCTGTGAGCTGATGCTTACAGCTGATGCTTACCAAGAGAGAGCTATCAGC 303
Db 127436 ACAAGCTGTGAGCTGATGCTTACAGCTGATGCTTACCAAGAGAGAGCTATCAGC 127495

OY 304 ATGAGCCCGTTCTCTGAGGAGCCAGATCATATATCTGTATAGCTTATCATCATAGAC 363
Db 127496 ATGAGCCCGTTCTCTGAGGAGCCAGATCATATATCTGTATAGCTTATCATCATAGAC 127555

OY 364 CTATTGAGAGGGTCTCTGATCAATGTTACTTCTTCTGTTGGTAAAGATTT 420
Db 127556 CTATTGAGAGGGTCTCTGATCAATGTTACTTCTTCTGTTGGTAAAGATTT 127612

RESULT 12
AF003280 147201 bp DNA linear PLN 17-APR-2002
LOCUS AP003280

DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
PAC clone: P0557A01.
ACCESSION AP003280
VERSION GI:16554287
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
clone: P0557A01.
Oryza sativa (japonica cultivar-group)
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzaceae; Oryza.
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0557A01
2 (bases 1 to 147201)
Published only in Database (2001)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Oct 31, 2001 this sequence version replaced gi:13027310.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db/) and the cDNA sequence database at
RGP. Protein homologues of the coding regions were searched against
NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from sp6 to T7 of the PAC clone.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rpg.dna.affrc.go.jp/GenomeSeq.html.

FEATURES
source
Location/Qualifiers
1..147201
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0557A01"
/join(4608..4682,4991..5422)
/gene="P0557A01.1"
/join(4608..4682,4991..5422)
/gene="P0557A01.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB89763.1"
/db_xref="GI:20160823"
/translation="MERGVAAARVGRSRGRPRPADGESFSGOGLVSTHGYMYT
RKHCRSHRSIOLLSAHHSILYRGFLSEGGWHLTRCIGRSPAKSALRSAGPNN
CIGYDIVGCVISWIALVALYLIPNLSTIYVNLVLAIRNSIGFSLYGCYRR
LNRRLRLD"
/join(7099..7303,8026..8447)
/gene="P0557A01.2"
/join(7099..7303,8026..8447)
/note="P0557A01.2"
/gene="contains Est C74750(E50835)
similar to Arabidopsis thaliana chromosome 3, At3g20570
unknown protein"
/codon_start=1
/protein_id="BAB89764.1"

OY 244 ACACTGTTGGAGTATGACCTTACCAAGTCTGATGCTTACAAAGACGATGCTATCAC 303
 DB 92824 ACACTGTTGGAGTATGACCTTACCAAGTCTGATGCTTACAAAGACGATGCTATCAC 92883
 OY 304 ATGAGCCGCTTCTGGGACCCAGATCATATATCTGTATGATGCTTATCATGAC 363
 DB 92884 ATGAGCCGCTTCTGGGACCCAGATCATATATCTGTATGATGCTTATCATGAC 92943
 OY 364 CTATTGGAAGAGGCTTCTGATGCTTACATGCTTATCTTACTGTTGGGATACGATTT 420
 DB 92944 CTATTGGAAGAGGCTTCTGATGCTTACATGCTTATCTTACTGTTGGGATACGATTT 93000

RESULT 13
 LOCUS AP003986 105649 bp DNA linear HTG 21-MAR-2002
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone OJ1057_A09, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AP003986
 VERSION AP003986.1 GI:15076789
 KEYWORDS HTG, HTGS, PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:OJ1057_A09.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC clone:OJ1057_A09
 JOURNAL Published Only in Database (2001)
 REFERENCE 2
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2001) Takuji Sasaki, National Institute of Agricultural Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nri.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..105649
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="OJ1057_A09"
 BASE COUNT 32062 a 21297 c 21329 g 30937 t 24 others
 ORIGIN

Query Match 91.3%; Score 383.4; DB 2; Length 105649;
 Best Local Similarity 95.0%; Pred. No. 1.8e-97;
 Matches 396; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 4 CGTTGGAGAAATCTTAATTCATGAGTGTAGGAGGAGCTTATGTCACCAACAGAA 63
 DB 13085 CGTTGGAGAAATCTTAATTCATGAGTGTAGGAGGAGCTTATGTCACCAACAGAA 13144
 OY 64 ACRAAGCAAGTGTGATTTAAAGCTGTGAAGGATTAATAAATGCTACTACAC 123
 DB 13445 ACRAAGCAAGTGTGATTTAAAGCTGTGAAGGATTAATAAATGCTACTACAC 13204

DB 13145 ACRAAGCAAGTGTGATTTAAAGCTGTGAAGGATTAATAAATGCTACTACAC 13204
 OY 124 CCGAGTACGAACCAAGATATGATATCTTGAGCATCTCGATACCTCGAGTC 183
 DB 13205 CCGAGTACGAACCAAGATATGATATCTTGAGCATCTCGATACCTCGAGTC 13264
 OY 184 GGGGTTCCGGCTGAAGAGAGAGAGCTGCAAGTACCTGCGAATCTTACTGCTACATG 243
 DB 13265 GGGGTTCCGGCTGAAGAGAGAGAGCTGCAAGTACCTGCGAATCTTACTGCTACATG 13324
 OY 244 ACACTGTTGGAGTATGACCTTACCAAGTCTGATGCTTACAAAGACGATGCTATCAC 303
 DB 13325 ACACTGTTGGAGTATGACCTTACCAAGTCTGATGCTTACAAAGACGATGCTATCAC 13384
 OY 304 ATGAGCCGCTTCTGGGACCCAGATCATATATCTGTATGATGCTTATCATGAC 363
 DB 13385 ATGAGCCGCTTCTGGGACCCAGATCATATATCTGTATGATGCTTATCATGAC 13444
 OY 364 CTATTGGAAGAGGCTTCTGATGCTTACATGCTTATCTTACTGTTGGGATACGATTT 420
 DB 13445 CTATTGGAAGAGGCTTCTGATGCTTACATGCTTATCTTACTGTTGGGATACGATTT 13501

RESULT 14
 LOCUS AP003623/c 139914 bp DNA linear HTG 21-MAR-2002
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0642B07, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AP003623
 VERSION AP003623.1 GI:14020961
 KEYWORDS HTG, HTGS, PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:P0642B07.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC clone:P0642B07
 JOURNAL Published Only in Database (2001)
 REFERENCE 2
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2001) Takuji Sasaki, National Institute of Agricultural Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nri.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..139914
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="P0642B07"
 BASE COUNT 41395 a 28837 c 28251 g 41381 t 50 others
 ORIGIN

Query Match 91.3%; Score 383.4; DB 2; Length 139914;
 Best Local Similarity 95.0%; Pred. No. 1.8e-97;
 Matches 396; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 12:46:46 ; Search time 3925 Seconds

(without alignments)
240.978 Million cell updates/sec

Title: US-09-696-664A-3

Perfect score: 420

Sequence: 1 gcgcgtgtgagattctta.....tcattgttggaacgatttt 420

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Matched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	91.7	134525	11	AAQ04525
2	340.8	81.1	4260	22	AAFB8278
3	177	42.1	1416	16	AAO87822
4	177	42.1	4234	16	AAO87819
5	133.2	31.7	2222	13	AAQ22147
6	129.4	30.8	1419	24	ABQ90432
7	80.4	19.1	1953	21	AAZ61381
8	80	19.0	195	21	AAZ61365
9	64	15.2	195	21	AAZ61367

10	44.2	10.5	159	21	AAZ61366
11	41	9.8	1416	20	AAZ21432
12	41	9.8	1417	18	AAZ85196
13	39.2	9.3	278	20	AAZ21420
14	38	9.0	6433	20	AAZ13041
15	36.6	8.7	349980	22	AAH41225
16	36.2	8.6	191	20	AAZ21410
17	35.6	8.5	2543	21	AAZ52774
18	35.6	8.5	3746	21	AAZ99512
19	35.6	8.5	3746	21	AAZ99527
20	35.6	8.5	3773	21	AAZ99494
21	35	8.3	4270	24	AB199885
22	34.8	8.3	22008	22	ABA15839
23	34.8	8.3	22008	22	AAZ36610
24	34.4	8.2	165	20	AAZ21409
25	34.4	8.2	465237	24	ABQ87681
26	34.4	8.2	465237	24	ABQ90193
27	34.2	8.1	3381	19	AAV41443
28	33.4	8.0	337	22	AAK67586
29	33.4	8.0	473	22	AAK56227
30	33.4	8.0	3595	23	ABJ12127
31	33	7.9	1827	24	ABQ70608
32	32.8	7.8	495	23	ABK42341
33	32.8	7.8	495	23	ABK42342
34	32.6	7.8	2373	22	AAH65995
35	32.6	7.8	349980	22	AAH68527
36	32.4	7.7	568	21	AAH44116
37	32.4	7.7	3191	22	AAZ58978
38	32.4	7.7	3363	22	AAZ45004
39	32.2	7.6	580073	18	AAZ58840
40	32	7.6	1693	21	AAZ59175
41	32	7.6	1696	24	ABO54262
42	32	7.6	2807	24	ABK48311
43	32	7.6	2992	24	ABN85281
44	32	7.6	5954	19	AAV59105
45	31.8	7.6	168	20	AAZ21407

ALIGNMENTS

RESULT 1	
ID	AAQ04525 standard; DNA: 134525 BP.
XX	AAQ04525;
XX	01-OCT-1990 (first entry)
XX	Total base sequence of rice plant chloroplast DNA.
XX	Chloroplast; true grass; rice plant; ss.
XX	Oryza sativa.
XX	JP02100682-A.
XX	12-APR-1990.
XX	07-OCT-1988; 88JP-0251967.
XX	07-OCT-1988; 88JP-0251967.
XX	07-OCT-1988; 88JP-0251967.
XX	(MITK) MITSUI TOANISU CHEM INC.
XX	WPI; 1990-159709/21.
XX	Chloroplast DNA of true grasses - used to produce various
XX	DNA base sequences by decomposition of rice plant DNA.
XX	Claim 1; Fig 1; 20pp; Japanese.
XX	The sequence is that of the whole of rice chloroplast DNA.

Nucleotide sequence
Regulatory region
Plastid targeting
5' and 3' regulatory
Enterococcus faecalis
Pyrococcus abyssi
Piryn promoter seg
Soybean putative c
DNA encoding a mal
DNA encoding a mal
DNA encoding a mal
Mouse ischaemic co
Human nervous syst
Human cardiovascular
Piryn promoter seg
Human oestrogen re
Human oestrogen re
M. catarrhalis str
Human immune/haema
Human immune/haema
Drosophila melanog
Listeria monocytog
Genomic sequence #
Genomic sequence #
C glutamicum codin
Human secreted exp
Human cytoskeletal
cDNA encoding nove
Mycoplasmata genital
Human secreted pro
Human ovarian anti
DNA encoding cap-b
Cell cycle regulat
Zebrafish differen
prrn/tbcl/Rubisco

CC	The methods comprise ligting an operon containing a promoter and 2-100
CC	genes of interest, to a vector and integrating the resulting recombinant
CC	vector into a plastid chromosome. The methods are useful for transforming
CC	a plant, belonging to any family such as Gramineae, Malvaceae,
CC	Brassicaceae, Compositae, Pedaliaceae, Oleaceae, Myrtaceae, Rosaceae,
CC	Nicotiana, Leguminosae, Palmae, Sterculiaceae or Rubiaceae, in particular
CC	Nicotiana tabacum of Solanaceae family, for producing polyester, which is
CC	copolymer of lower alkyl 3-hydroxyalkanoic acid. In the present invention
CC	to generate a plant which produces polyester. In order to ensure that
CC	homologous recombination is carried out in the plastid chromosome, the
CC	present sequence, the rpoL and ORF512 genes, which exist in the plastid
CC	were used.
XX	
SQ	Sequence 4260 BP; 1336 A; 675 C; 892 G; 1357 T; 0 other;
Query_Match	81.1%; Score 340.8; DB 22; Length 4260;
Best Local Similarity	88.7%; Pred. No. 2.4e-94;
Matches 369; Conservative	0; Mismatches 47; Indels 0; Gaps 0;
OY	5 GTTGTGAGAAATCTTAATCATGAGTTTAGGAGGAGGACTTATGTACCAACAACGAAA 64
Dd	
Db	186 GTTGAGAGATTCTTAATCATGAGTTTAGGAGGAGGATTATGTACCAACAACGAGA 245
OY	65 CTAAAGCAAGTGTGGATTAAAGCTGTGTGAAGGATTATAAATGACTTACTACACC 124
Dd	
Db	246 CTAAAGCAAGTGTGGATTCAAAAGCTGTGTTAAAGAGTACAAATTGACTTATATAC 305
OY	125 GCGAGTACCAAAACCAAGATGATGATATCTTGGCACATTCGAGTAACTCCTCAGACTG 184
Dd	
Db	306 CTGAGTACCAAAACCAAGATGATGATATCTTGGCACATTCGAGTAACTCCTCAGACTG 365
OY	185 GGGTTCGCCCTTGAAGAAGCAGGAGCTGCAGTACGTCCGCGAATCTTACTGTGTAATGSA 244
Dd	
Db	366 GAGTTCCACCTGAAGAAGAGGAGGGCGCGGSGTAGCTGCCAATCTTCTGATGTAATGSA 425
OY	245 CAACGTITGGACTATGACTTACCAAGTCTTGATGCTGTACAAAGAGAGATCTATACAA 304
Dd	
Db	426 CAACGTATGAGCCACATGAGACTTACCAAGCTTGATGCTGTACAAAGAGAGATCTATAC 485
OY	305 TCGAGCCGCTTCCTGGGGGACCCAGATCAATATATCTGTATGTAGTACTTATCATTAAGAC 364
Dd	
Db	486 TCGAGCCGCTTCCTGGGGGAAAAAGATCAATATATGTCTTATGTAGTACTTATACCTTTAGACC 545
OY	365 TATTTGAAGAGGGTCTTGTACTAACATGTTTACTTTCATTTGGGTGAACGATATT 420
Dd	
Db	546 TTTTGAAGAAGGTTCTGTACCAACATGTTTACTTTCATTTGTAGTAAAGATATT 601
RESULT 3	
AQ087822	
ID	AQ087822 standard; DNA; 1416 BP.
XX	
AC	AQ087822:
XX	
DT	14-DEC-1995 (first entry)
XX	
DE	Agmenellum quadruplicatum PR-6 rubisco large subunit coding region.
XX	
KW	Rubisco; ribulose 1,5-bisphosphate carboxylase/oxygenase;
KW	large subunit; carbon dioxide fixation; Synechococcus; ds.
XX	
OS	Agmenellum quadruplicatum PR-6.
PH	
FH	
FT	Key Location/Qualifiers
FT	CDS 1..1416
FT	/tag= a
XX	/product= large_subunit
PN	JF07079782-A.
XD	
DB	28-MAR-1995.

PF 18-JUN-1993; 93JP-0184304.
 XX 18-JUN-1993; 93JP-0184304.
 PR (KANT) KANSAI DENRYOKU KK.
 XX WPI, 1995-157852/21.
 DR P-PSDB; AAR72392.
 XX Promoter sequence of *Synechococcus PCC7002*-originated rubisco gene -
 PT used in a vector to produce *Cyano:bacterium* with improved carbon
 PT di:oxide fixation
 PS Claim 7; Page 17-18; 28pp; Japanese.
 XX A sequence including the ORFs coding for the large and small
 CC subunits of the rubisco enzyme has been isolated from *Agmenellum*
 CC quadruplicatum PR-6 (*Synechococcus* sp. strain PCC7002). The
 CC region coding for the rubisco large subunit is claimed; the enzyme
 CC is involved in carbon dioxide fixation.

Query Match 42.1%; Score 177; DB 16; Length 1416;
 Best Local Similarity 69.6%; Pred. No. 3.9e-44;
 Matches 240; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 75 TGTGATTAAAGCTGGTGTAAAGATTAATAATTGACTTACACCCCGAGTACGA 134
 DB 18 TGTGATTAAAGCTGGTGTAAAGATTAATAATTGACTTACACCCCGAGTACGA 77
 QY 135 AACCAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 194
 DB 78 CCCGAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 137
 QY 195 TGAAGAGCAGGAGCTGAGTACGAGTATCTTCTACTGAGTACGAGTATCTTGG 254
 DB 138 CGAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 197
 QY 255 GACTGATGAGTACGAGTATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 314
 DB 198 GACGATGAGTACGAGTATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 257
 QY 315 TCCTGGGAGCCAGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 374
 DB 258 TCCTGGGAGCCAGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 317
 QY 375 GGTTCCTGACTACATGTTTACTTCAATTTGTTGGTAACTGATTT 419
 DB 318 AGGTTCCTGACTACATGTTTACTTCAATTTGTTGGTAACTGATTT 362

RESULT 4
 AA087819
 ID AA087819 standard; DNA; 4234 BP.
 AC AA087819;
 XX 14-DEC-1995 (first entry)
 XX *Agmenellum quadruplicatum* PR-6 rubisco gene.
 DE *Agmenellum quadruplicatum* PR-6 rubisco gene.
 XX Rubisco: ribulose 1,5-bisphosphate carboxylase/oxygenase; promoter;
 KW carbon dioxide fixation; *Synechococcus* sp.; ds.
 XX *Agmenellum quadruplicatum* PR-6.
 OS
 XX Key location/Qualifiers
 PH 45..344
 FT CDS
 FT /*tag= a
 FT /label= ORF3
 FT /note= "claim 5, encodes AAR72391"
 FT promoter 659..962

FT /*tag= b
 FT /note= "claim 1"
 FT 963..2378
 FT /*tag= c
 FT /product= large_subunit
 FT /note= "claim 7, encodes AAR72392"
 FT 2447..2851
 FT /*tag= d
 FT /label= ORF2
 FT /note= "claim 9, encodes AAR72393"
 FT 2882..3217
 FT /*tag= e
 FT /product= small_subunit
 FT /note= "claim 11, encodes AAR72394"
 FT 3379..4005
 FT /*tag= f
 FT /label= ORF1
 FT /note= "claim 13, encodes AAR72395"

JP07079782-A.
 XX 28-MAR-1995.
 XX 18-JUN-1993; 93JP-0184304.
 XX 18-JUN-1993; 93JP-0184304.
 PR (KANT) KANSAI DENRYOKU KK.
 XX WPI, 1995-157852/21.
 DR P-PSDB; AAR72391, AAR72392, AAR72393, AAR72394, AAR72395.
 XX Promoter sequence of *Synechococcus PCC7002*-originated rubisco gene -
 PT used in a vector to produce *Cyano:bacterium* with improved carbon
 PT di:oxide fixation
 PS Example 7; Page 9-14; 28pp; Japanese.
 XX A sequence including the ORFs coding for the large and small
 CC subunits of the rubisco enzyme has been isolated from *Agmenellum*
 CC quadruplicatum PR-6 (*Synechococcus* sp. strain PCC7002). The
 CC promoter region and various coding regions are each claimed (see
 CC Features Table). The rubisco enzyme is involved in carbon dioxide
 CC fixation.

Query Match 42.1%; Score 177; DB 16; Length 4234;
 Best Local Similarity 69.6%; Pred. No. 6.2e-44;
 Matches 240; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 75 TGTGATTAAAGCTGGTGTAAAGATTAATAATTGACTTACACCCCGAGTACGA 134
 DB 980 TGTGATTAAAGCTGGTGTAAAGATTAATAATTGACTTACACCCCGAGTACGA 1039
 QY 135 AACCAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 194
 DB 1040 CCCGAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 1099
 QY 195 TGAAGAGCAGGAGCTGAGTACGAGTATCTTCTACTGAGTACGAGTATCTTGG 254
 DB 1100 CGAAGATACGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 1159
 QY 255 GACTGATGAGTACGAGTATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 314
 DB 1160 GACGATGAGTACGAGTATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 1219
 QY 315 TCCTGGGAGCCAGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 374
 DB 1220 TCCTGGGAGCCAGATATCTTGGACAGATTCGAGTACTCTCCAGCTCCGGCTCCGCC 1279
 QY 375 GGTTCCTGACTACATGTTTACTTCAATTTGTTGGTAACTGATTT 419

DB 1280 AGGTCGTGTAACCAACGTTTGGACTTCCTGTTGTGTAAGTATT 1324

RESULT 5
AAQ22147
ID AAQ22147 standard; DNA; 2222 BP.
XX
AC AAQ22147;
XX
DT 30-JUN-1992 (first entry)
XX
DE RubPCase gene.
XX
KW D-ribulose-1,5-bisphosphate carboxylase; Rubisco; bioreactor; carbon;
XX dioxide absorption; large subunit; small subunit; ss.
XX
OS Thiobacillus ferrooxidans.
XX
FH
FT Key Location/Qualifiers
FT CDS 181..1603
FT /tag= a
FT /product= large_subunit
FT 1693..2026
FT /tag= b
FT /product= small_subunit
XX
XX JP04030792-A.
XX
XX 03-FEB-1992.
XX
XX 29-MAY-1990; 90JP-0139300.
XX
XX 29-MAY-1990; 90JP-0139300.
XX
XX (DOWA) DOWA MINING CO LTD.
XX (AKIT) AKITA KEN.
XX
XX WPI: 1992-085966/11.
XX P-PSDB; AAR21977; AAR23903.
XX
XX RubPCase gene from Thiobacillus ferrooxidans - used to produce
XX RubPCase for CO2 absorption bioreactor
XX
XX Claim 1; Page 1; 16pp; Japanese.
XX
XX The D-ribulose-1,5-bisphosphate carboxylase (RubPCase) gene was
XX isolated from Thiobacillus ferrooxidans. The genomic DNA of T.
XX ferrooxidans was recovered from a 9K medium culture incubated with
XX protease K. A RubPCase gene fragment from Chromatium vinosum was
XX used as a probe in Southern hybridisation to genomic DNA. The T.
XX ferrooxidans Pel strain RubPCase gene was cloned and a deletion
XX plasmid constructed. The nucleotide and corresp. amino acid
XX sequences of RubPCase were determined. The gene encodes both the
XX large and small RubPCase subunits, which together form a complex of
XX 8 large and 8 small subunits designated Rubisco (EC 4.1.1.39).
XX An E. coli plasmid for expressing the RubPCase gene was constructed
XX and the RubPCase activity in E. coli determined. RubPCase produced
XX by E. coli or Thiobacillus can be used in a bioreactor of carbon
XX dioxide absorption.
XX
XX Sequence 2222 BP; 496 A; 623 C; 657 G; 446 T; 0 other;
XX
XX Query Match 31.7%; Score 133.2; DB 13; Length 2222;
XX Best Local Similarity 62.1%; Pred. No. 1.4e-30;
XX Matches 210; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

DB 196 TATGAGCCGGGTAAAGCACTACCGGACGATGATGGCCCCGAGTACGCTTTG 255

QY 142 GATGACTGATCTTGGCAGCATTCGAGTAACCTCCTAGTCCGGGTTCCGCTGAGAA 201

DB 256 GATTCGATATCCCTAGCCTGTTCAAGATTACCGCCGACAGGCGGTGATGAGGAA 315

QY 202 GCAGAGAGCTGCACTAGCTCGGAATCTTCTACTGCTAGACAGACTGTTGACTGAT 261

DB 316 GCCGCCGCCGCGGTAGCCGCCGAGTCTTCTACCGGCACATGACACGAGTGTGACCT 375

QY 262 GGAATTACCAAGTCTGATGCTTACAAAGACAGATGATCATCATGAGCCGTTCCCTGG 321

DB 376 TTGTGACCGACATGATGATTTACTACAAAGCCGCCGCTATGCAATGAAAGATGCGCCGG 435

QY 322 GACCCAGATCAATATATGTTATGATGCTTATCATATGACCTATTGTAAGAGGTTCT 381

DB 436 GACGACACCTGCTTTTACGCTTGTGCTTGCATGCCATTCATTTGTTAAGAGGCTCC 495

QY 382 TGTACTAACATGTTTACTTTCATGCTTGTGCTGCTGATTT 419

DB 496 GTGTTAACGCTTCTACCTCCTGCTGCTGTAAGCTGT 533

RESULT 6
ABQ90432
ID ABQ90432 standard; DNA; 1419 BP.
XX
AC ABQ90432;
XX
XX 01-OCT-2002 (first entry)
XX
XX M. capsulatus gene #417 for DNA array.
XX
XX Micro array; gene; ds; differential expression; gene expression.
XX
XX Methylococcus capsulatus.
XX
XX W0200255655-A2.
XX
XX 18-JUL-2002.
XX
XX 14-JAN-2002; 2002WO-NO00019.
XX
XX 12-JAN-2001; 2001NO-0000235.
XX 12-JAN-2001; 2001NO-0000239.
XX
XX (UNTF-) UNITEB STIFTELSEN UNIV BERGEN.
XX (TIGR-) TIGR.
XX
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
XX Lillehaug JR, Lossius I, Eisein JA, Frazer CM, Durkin AS;
XX Salzberg SL;
XX
XX WPI: 2002-557818/59.
XX
XX Novel DNA array useful for determining differential expression of
XX Methylococcus capsulatus genes, comprises polynucleotides or
XX oligonucleotides representative for a selective number of Methylococcus
XX capsulatus genes -
XX
XX Claim 13; Page 248; 678pp; English.
XX
XX The invention relates to a novel DNA array giving a representation of a
XX number of Methylococcus capsulatus genes. The method of the invention is
XX useful for determination of the differential expression of the genes of
XX M. capsulatus, and for studying gene expression on a genomic scale and in
XX ABQ90016-ABQ91855 represent M. capsulatus genes. The sequences shown in
XX invention.
XX
XX Sequence 1419 BP; 285 A; 493 C; 430 G; 211 T; 0 other;
XX
XX Query Match 30.8%; Score 129.4; DB 24; Length 1419;
XX Best Local Similarity 61.4%; Pred. No. 1.7e-29;
XX Matches 208; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 81 ATTAAAGCTGCTTAAAGATTAATTAATGACTTATACACCCGAGTACCAACCA 140

Db 15 ATACACGGGGCGGCAAGATACCCGGAACCTACTGGACCCGGAACCTACTCCCGC 74
 Qy 141 GGATACGATATCTTGGCAGCATTCGAGTAACTCTCAGCTCGGGGTTCGCGTGAAGA 200
 Db 75 CGACACCGATCTGCTGCTTCAAGATCACCCCGGGGTGCGGCGGCGAAGA 134
 Qy 201 AGCAGAGCTGATGCTGCGGAATCTTCTACGTGATACAGCAACTGTTGGACTGA 260
 Db 135 AGCGCCCGCCCGTGGCGGGGATGTCGACCGGACCGGACCGCTCTGGACCGA 194
 Qy 261 TGAACCTACAGTCTGATCTGATCAAAAGACATGCTATCATCGACCGCGTTCGCG 320
 Db 195 CCTGTGACCCACCTCGCATTTACAAAGGCGCCGCTACCGGATCGAGACGTGCGCG 254
 Qy 321 GGACCCAGCATATATCTGATGATGAGCTTATCATGACCTATTGGAAGAGGGTTC 380
 Db 255 CCAGGACGACAGATTCCTACGCGCTTACGCGCTATCCATCGACCTGTTGAGAGGGGCTC 314
 Db 381 TTGCTACTACATGTTTACTTTCATTCATTCGCGTAACTGAT 419
 Db 315 CGTGTCAACGTTGACCTCGTGTGGGCAACGTGTT 353

RESULT 7
 AAZ61381
 ID AAZ61381 standard; DNA; 1953 BP.
 XX AAZ61381;
 AC
 XX 19-JUN-2000 (first entry)
 DT
 XX Nucleotide sequence of DNA construct FLARE16-S1.
 DE
 XX Green fluorescent protein; GFP; aadA; aptB; protein expression; vaccine;
 KW haemoglobin; enzyme; psbA; ss.
 KM
 XX Synthetic.
 OS unidentified.
 OS Aequorea victoria.
 OS
 XX Key Location/Qualifiers
 FH 7..189
 FT /tag- a
 FT /note- "atpB downstream box"
 FT 196..978
 FT /tag- b
 FT /note- "aadA sequence"
 FT 1027..1747
 FT /tag- c
 FT /note- "green fluorescent protein region"
 FT 1758..1946
 FT /tag- d
 FT /note- "psbA region"
 XX
 XX WO200007431-A1.
 XX 17-FEB-2000.
 PD
 XX 03-AUG-1999; 99WO-US17806.
 PF
 XX 03-AUG-1998; 98US-0095163.
 PR 03-AUG-1998; 98US-0095167.
 PR 15-DEC-1998; 98US-0112257.
 PR 29-APR-1999; 99US-0131611.
 PR 11-JUN-1999; 99US-0138764.
 XX
 XX (RUTF) UNIV RUTGERS STATE NEW JERSEY.
 XX
 XX Maliga P, Kuroda H, Khan MS;
 XX WPI; 2000-205525/18.
 XX
 XX New recombinant DNA constructs, for expressing high levels of

PT heterologous protein in plastids of higher plants, includes promoter, a
 PS leader sequence and a downstream box element -
 XX Disclosure; Fig 29; 164pp; English.
 XX
 CC The present sequence represents a DNA construct of the invention.
 CC The specification describes recombinant DNA constructs for expressing
 CC heterologous proteins in the plastids of higher plants. The DNA
 CC constructs comprise a 5' regulatory region which includes a promoter
 CC element, a leader sequence and a downstream box element operably linked
 CC to a coding region of the heterologous protein. The chimeric regulatory
 CC region enhances translational efficiency of an mRNA molecule encoded by
 CC the DNA construct. The DNA constructs are used for producing transformed
 CC monocot and dicot plants having high levels of heterologous protein
 CC expression. They can be used to drive expression of proteins with
 CC agronomic, industrial or pharmaceutical importance, including production
 CC of vaccines, healthcare products like human haemoglobin, industrial or
 CC household enzymes. Plants which can be transformed with the constructs
 CC of the invention include maize, millet, sorghum, sugar cane, rice,
 CC wheat, barley, oat, rye or turf grass.
 CC
 XX Sequence 1953 BP; 546 A; 422 C; 487 G; 498 T; 0 other;

Query Match 19.1%; Score 80.4; DB 21; Length 1953;
 Best Local Similarity 93.3%; Pred. No. 2.4e-14;
 Matches 84; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 5 GTTGTGAGAAATCTTAATCATGAGTTGAGGAGGACTTATGTCACCAACAGAAA 64
 Db 107 GTTGTGAAATCTTAATCATGAGTTGAGGAGGACTTATGTCACCAACAGAAA 166
 Qy 65 CTAAAGCAAGTGTGATTTAAAGCTGGTG 94
 Db 167 CTAAAGCAAGTGTGATTTAAAGCTAGCG 196

RESULT 8
 AAZ61365
 ID AAZ61365 standard; DNA; 195 BP.
 XX AAZ61365;
 AC
 XX 19-JUN-2000 (first entry)
 DT
 XX Nucleotide sequence of chimeric promoter Prrn1bcl-DBwt.
 DE
 XX Chimeric promoter; plastid RNA operon omega-type promoter; haemoglobin;
 KW Prrn promoter; rbcL; protein expression; vaccine; enzyme; ss.
 KM
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH 7..89
 FT /tag- a
 FT /note- "Prrn plastid promoter"
 FT 138..142
 FT /tag- b
 FT /note- "Shine-Dalgarno sequence"
 XX
 XX WO200007431-A1.
 XX 17-FEB-2000.
 PD
 XX 03-AUG-1999; 99WO-US17806.
 PF
 XX 03-AUG-1998; 98US-0095163.
 PR 03-AUG-1998; 98US-0095167.
 PR 15-DEC-1998; 98US-0112257.
 PR 29-APR-1999; 99US-0131611.
 PR 11-JUN-1999; 99US-0138764.
 XX
 XX (RUTF) UNIV RUTGERS STATE NEW JERSEY.

PI Maliga P, Kuroda H, Khan MS;
 XX WPI, 2000-205525/18.
 XX
 XX
 PT New recombinant DNA constructs, for expressing high levels of
 PT heterologous protein in plastids of higher plants, includes promoter, a
 PT leader sequence and a downstream box element -
 XX
 XX Claim 3, Fig 3B, 164pp; English.
 XX
 CC The present sequence represents a chimeric plastid RNA operon
 CC omega-type (Prn) promoter with rbcL translation control sequences and
 CC a wild type downstream box. The chimeric promoter is used, as a 5'
 CC regulatory sequence, to produce recombinant DNA constructs for expressing
 CC heterologous proteins in the plastids of higher plants. The DNA
 CC constructs comprise a 5' regulatory region which includes a promoter
 CC element, a leader sequence and a downstream box element operably linked
 CC to a coding region of the heterologous protein. The chimeric regulatory
 CC region enhances translational efficiency of an mRNA molecule encoded by
 CC the DNA construct. The DNA constructs are used for producing transformed
 CC monocot and dicot plants having high levels of heterologous protein
 CC expression. They can be used to drive expression of proteins with
 CC agronomic, industrial or pharmaceutical importance, including production
 CC of vaccines, healthcare products like human haemoglobin, industrial or
 CC household enzymes. Plants which can be transformed with the constructs
 CC of the invention include maize, millet, sorghum, sugar cane, rice,
 CC wheat, barley, oat, rye or turf grass.
 CC
 SQ Sequence 195 BP; 51 A; 33 C; 60 G; 51 T; 0 other;
 XX
 XX Query Match 19.0%; Score 80; DB 21; Length 195;
 XX Best local Similarity 94.38; Pred. No. 1.2e-14;
 XX Matches 83; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 5 GTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGTATGTCACCAACAGAAA 64
 DB 107 GTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGTATGTCACCAACAGAAA 166
 OY 65 CTAAGCAAGTGTGATTTAAAGCTGG 92
 DB 167 CTAAGCAAGTGTGATTTAAAGCTGG 194
 XX
 XX RESULT 9
 XX ID AA261367 standard; DNA; 195 BP.
 XX AC AA261367;
 XX DT 19-JUN-2000 (first entry)
 XX
 XX Nucleotide sequence of chimeric promoter PrnLrbcL+DBm.
 KW Chimeric promoter; plastid RNA operon omega-type promoter; haemoglobin;
 KW Prn promoter; rbcL; protein expression; vaccine; enzyme; ss.
 XX
 XX Synthetic.
 OS
 XX
 XX Key Location/Qualifiers
 XX FH 7..89
 XX FT promoter /note= "Prn plastid promoter"
 XX FT 138..142
 XX FT misc_signal /note= "Shine-Dalgarno sequence"
 XX FT 153..175
 XX FT /tag= c
 XX
 XX WO200007431-A1.
 XX 17-FEB-2000.
 XX 03-AUG-1999; 99WO-US17806.

XX 03-AUG-1998; 98US-0095163.
 PR 03-AUG-1998; 98US-0095167.
 PR 15-DEC-1998; 98US-0112257.
 PR 29-APR-1999; 99US-0131611.
 PR 11-JUN-1999; 99US-0138764.
 XX
 XX (RUTE) UNIV RUTGERS STATE NEW JERSEY.
 XX
 XX Maliga P, Kuroda H, Khan MS;
 XX WPI, 2000-205525/18.
 XX
 XX
 PT New recombinant DNA constructs, for expressing high levels of
 PT heterologous protein in plastids of higher plants, includes promoter, a
 PT leader sequence and a downstream box element -
 XX
 XX Claim 3, Fig 3B, 164pp; English.
 XX
 CC The present sequence represents a chimeric plastid RNA operon
 CC omega-type (Prn) promoter with rbcL translation control sequences and
 CC a mutated downstream box. The chimeric promoter is used, as a 5'
 CC regulatory sequence, to produce recombinant DNA constructs for expressing
 CC heterologous proteins in the plastids of higher plants. The DNA
 CC constructs comprise a 5' regulatory region which includes a promoter
 CC element, a leader sequence and a downstream box element operably linked
 CC to a coding region of the heterologous protein. The chimeric regulatory
 CC region enhances translational efficiency of an mRNA molecule encoded by
 CC the DNA construct. The DNA constructs are used for producing transformed
 CC monocot and dicot plants having high levels of heterologous protein
 CC expression. They can be used to drive expression of proteins with
 CC agronomic, industrial or pharmaceutical importance, including production
 CC of vaccines, healthcare products like human haemoglobin, industrial or
 CC household enzymes. Plants which can be transformed with the constructs
 CC of the invention include maize, millet, sorghum, sugar cane, rice,
 CC wheat, barley, oat, rye or turf grass.
 CC
 SQ Sequence 195 BP; 51 A; 34 C; 60 G; 47 T; 3 U; 0 other;
 XX
 XX Query Match 15.2%; Score 64; DB 21; Length 195;
 XX Best local Similarity 83.0%; Pred. No. 1e-09; 15; Indels 0; Gaps 0;
 XX Matches 73; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 OY 5 GTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGTATGTCACCAACAGAAA 64
 DB 107 GTTGTGAGAAATCTTAATTCATGAGTTAGGAGGAGTATGAGUCUCAGACAGAAA 166
 OY 65 CTAAGCAAGTGTGATTTAAAGCTGG 92
 DB 167 CAAAGCCUCAGTAGGATTCAAAGCTAG 194
 XX
 XX RESULT 10
 XX ID AA261366 standard; DNA; 159 BP.
 XX AC AA261366;
 XX DT 19-JUN-2000 (first entry)
 XX
 XX Nucleotide sequence of chimeric promoter PrnLrbcL-DB.
 KW Chimeric promoter; plastid RNA operon omega-type promoter; haemoglobin;
 KW Prn promoter; rbcL; protein expression; vaccine; enzyme; ss.
 XX
 XX Synthetic.
 OS
 XX
 XX Key Location/Qualifiers
 XX FH 7..89
 XX FT promoter /note= "Prn plastid promoter"
 XX FT 138..142
 XX FT /tag= b
 XX
 XX misc_signal

/note= "Shine-Dalgarno sequence"

```

FT XX      /note= "Shine-Dalgarno sequence"
XX      WO200007431-A1.
XX      17-FEB-2000.
XX      03-AUG-1999; 99WO-US17806.
XX      03-AUG-1998; 98US-0095163.
XX      03-AUG-1998; 98US-0095167.
XX      15-DEC-1998; 98US-0112257.
XX      29-APR-1999; 99US-0131611.
XX      11-JUN-1999; 99US-0138764.
XX      (RUTE ) UNIV RUTGERS STATE NEW JERSEY.
XX      Maliya P, Kuroda H, Khan MS;
XX      WPI; 2000-205525/18.
XX      Claim 3; Fig 3B; 164pp; English.
XX      The present sequence represents a chimeric plasmid RNA operon
XX      omega-type (Prm) promoter with rbcL translation control sequences,
XX      without a downstream box. The chimeric promoter is used, as a 5'
XX      regulatory sequence, to produce recombinant DNA constructs for expressing
XX      heterologous proteins in the plastids of higher plants. The DNA
XX      constructs comprise a 5' regulatory region which includes a promoter
XX      element, a leader sequence and a downstream box element operably linked
XX      to a coding region of the heterologous protein. The chimeric regulatory
XX      region enhances translational efficiency of an mRNA molecule encoded by
XX      the DNA construct. The DNA constructs are used for producing transformed
XX      monocot and dicot plants having high levels of heterologous protein
XX      expression. They can be used to drive expression of proteins with
XX      agronomic, industrial or pharmaceutical importance, including production
XX      of vaccines, healthcare products like human hemoglobin, industrial or
XX      household enzymes. Plants which can be transformed with the constructs
XX      of the invention include maize, millet, sorghum, sugar cane, rice,
XX      wheat, barley, oat, rye or turf grass.
XX      Sequence 159 BP; 35 A; 26 C; 53 G; 45 T; 0 other;
XX      Query Match 10.5%; Score 44.2; DB 21; Length 159;
XX      Best Local Similarity 93.9%; Pred. No. 0.0012;
XX      Matches 46; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 5 GTTGTGAGAAATCTTAATTCATGAGTGTAGGAGGACTTATGTACAC 53
DB 107 GTTGTGAGAAATCTTAATTCATGAGTGTAGGAGGACTTATGTACAG 155
XX
RESULT 11
XX AAX21432
XX ID AAX21432 standard; DNA; 1416 BP.
XX
XX AAX21432;
XX
XX 21-MAY-1999 (first entry)
XX
XX Regulatory region Prm(L)/rbcL(s)/Kan/TpsbA(L).
XX
XX Construct; marker; antibiotic resistance; regulatory sequence; promoter;
XX stabilizing sequence; plastid; plant; ss.
XX
XX Synthetic.
XX
XX OS US5877402-A.
XX PN
XX PR 02-MAR-1999.
XX PD

```

```

XX      31-JAN-1994; 94US-0189256.
XX      31-JAN-1994; 94US-0189256.
XX      01-MAY-1990; 90US-0518763.
XX      25-AUG-1993; 93US-0111398.
XX      (RUTE ) UNIV RUTGERS STATE NEW JERSEY.
XX      Allison LA, Carrier H, Kanevski I, Maliya P, Maliya ZS;
XX      Staub JM, Zoubenko O;
XX      WPI; 1999-189744/16.
XX      New DNA construct containing an antibiotic resistance marker
XX      useful for stably transforming the plastids of multicellular plants
XX      Disclosure; Fig 28A; 101pp; English.
XX      The invention relates to a DNA construct, containing a non-lethal stable
XX      marker (e.g antibiotic resistance) under the control of a 5' regulatory
XX      sequence and a 3' stabilizing sequence, for stably transforming the
XX      plastids of multicellular plants and allowing expression of heterologous
XX      proteins especially non-native plastid or plant proteins. This sequence
XX      corresponds to the regulatory region Prm(L)/rbcL(s)/Kan/TpsbA(L) from
XX      the plasmid pRNH7.
XX      Sequence 1416 BP; 312 A; 322 C; 379 G; 403 T; 0 other;
XX      Query Match 9.8%; Score 41; DB 20; Length 1416;
XX      Best Local Similarity 72.6%; Pred. No. 0.028;
XX      Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
OY 28 AGTGTAGGAGGAGTATGTCACCAACAACTAAAGCAAGTGTGATTAA 87
DB 139 AGTGTAGGAGGAGTATGTCACCAACAACTAAAGCAAGTGTGATTAA 198
XX
OY 88 GCTGTGTAAAG 100
DB 199 GCAGGTTCTCCG 211
XX
RESULT 12
XX AAT85196
XX ID AAT85196 standard; DNA; 1417 BP.
XX
XX AAT85196;
XX
XX 10-MAR-1998 (first entry)
XX
XX Plasmid targeting region of plasmid pGS85A.
XX
XX Plasmid; transformation; transplastomic plant; transgenic plant;
XX Brassica; cruciferous plant; vector; plasmid pGS85A;
XX kanamycin resistance; neomycin phosphotransferase; neo gene;
XX selectable marker; ss.
XX
XX Chimeric - Arabidopsis thaliana.
XX Chimeric - Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 158..970
XX /*tag= a
XX /note= "kanamycin resistance gene"
XX
XX WO9732977-A1.
XX
XX 12-SEP-1997.
XX
XX 06-MAR-1997; 97WO-US03444.
XX
XX 06-MAR-1996; 96US-0012916.
XX
XX

```

31-JAN-1994; 94US-0189256.

CC A computer readable medium has been developed which has recorded on it

982 nucleotide sequences isolated from the *Enterococcus faecalis* genome
MAX12938 to MAX13919 represent these nucleotide sequences which are
primary nucleotide sequences, also known as contigs. The computer-based
system can identify fragments of the *Enterococcus faecalis* genome with
commercial importance. The products can be used to detect the presence
of *Enterococcus faecalis* in samples. They can also be used for
diagnosing *Enterococcal* infection in an animal and monitoring
progression of disease, and for identifying agents which can be used to
modulate the growth or pathogenicity of *Enterococcus faecalis*, or
another related organism, in vivo or in vitro. In particular the
polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
can be used in vaccines to prevent or attenuate an *Enterococcal*
infection.

SQ Sequence 6433 BP; 1816 A; 1501 C; 1085 G; 2022 T; 9 other;

Query Match	9.0%	Score 38;	DB 20;	Length 6433;
Best Local Similarity	46.0%	Pred. No. 0.43;		
Matches 128; Conservative	0;	Mismatches 150;	Indels 0;	Gaps 0;

QY	3	GCCTTGGAGAAATTCCTTAATTCATGAGTTGTAGGAGGAGCACTTAATGTCCACCAAAACAGA	62
Db	3674	GGGCTTTAAATATTCATCCAAAGGCTCATTTTGTTCAGCCCAATTTGAAGCAACCCGGCGATG	373/33
QY	63	AACTAAACGAAGTTGGATTTAAAGCTGGTGTAAAGATTAAATTGACTTACTACAC	122
Db	3734	CTCTTAACTATGACACCTTTACAGTTGGTGGCAATGATCATATCATCTCGGTGAC	379/33
QY	123	CGCGAGTACGAACCAACGAAGTACTGTAATCTTGGAGACATCCGATACCTCTGACT	182
Db	3794	CCCGAATATCCCACTATATGATACGTAAAGAATGATCATTTGTCCATTAAACTTTGGCT	385/5
QY	183	CGGGGTCCCGCTGAGAAGACGAGCGATGAGTACGTGCGGAATCTTACTGTGATACG	242
Db	3854	ACTAATTCCTCTTGGTAACCTGTCATGCCAGTAGTAAACAACTTCCTCTACTACAT	391/33
QY	243	GACAACTGTTTGGACTATGACTTACCACTCTTGATC	280
Db	3914	GCCTTCTGGACCAACGCTTTTCCCTCAAACTACTGTTTC	395/1

RESULT 15
AAH41225/c
ID AAH41225 standard; DNA; 349980 BP

AAH41225;

29-OCT-2001 (first entry)

DE pyrococcus abyssi genomic fragment #4.

Hyperthermophilic archaeon; hyperthermophilic protein; ds-

05 *Pyrococcus abyssi*.

	Key	Location/Qualifiers
FH	misc_feature	1..49980
FT		

FT	/note="this sequence overlaps with the 3' end of
FT	AHA41224"
FT	300001..349980
FT	/*tag= b
FT	/note="this sequence overlaps with the 5'end of
FT	AHA41226"

PN FR2792651-A1.

PD 27-OCT-2000.

PF 21-APR-1999; 99FR-0005034.

PR 21-APR-1999; 99FR-0005034.

XX

PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Hellig R;

PT New nucleotide sequences isolated from *Pyrococcus abyssi* encode
PT proteins useful in industry -

PS Claim 1; Page 429-524; 1657pp; French

The present invention relates to the genomic sequence of *P. abyssi* and *P. abyssi* proteins (see AA096053-AA096842). *P. abyssi* is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present sequence is a fragment of the genomic sequence of *P. abyssi*. The 5' end of this sequence overlaps with the 3' end of AA041224 and the 3' end of this sequence overlaps with the 5' end of AA041226. The proteins of the present invention have various commercial industrial uses since the proteins are stable at very high temperatures, some up to 110 degrees centigrade. In the context of family 28, WO2000/055062, which

CC Note: This patent is in the same patent family as WO200605062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAB75903-AAB75920 and AAG66436.

Sequence 349980 BP; 99421 A; 76616 C; 77444 G; 96499 T; 0 other;

Query Match	8.78;	Score 36.6;	DB 22;	Length 349980;
Best Local Similarity	60.6%;	Pred. No. 6.1;		
Matches 60;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0

QY 321 GGCCCCAGTCAATTAATCTGTTATGAGCTTATCATTAACCATTTGGAAGGGCTTC 380
 Db 268563 GCACGCGCGAGGGTTATATAGCAAAATATGCTTAACCCCTAACCCATTGGAAGGGAG 268504
 QY 381 TTGTACTAACAAGTTTACTTTCATCTGTGGTCAAGCTATT 419
 Db 268503 CTTAGTTCAGCTGTTTAGTGCCTATAGCTCGGACGACTATT 268465

Search completed: April 16, 2003, 02:32:58
Job time : 4060 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 11:36:59 ; Search time 91 Seconds

(Without alignments)
1415.431 Million cell updates/sec

Title: US-09-696-664A-3

Perfect score: 420

Sequence: 1 ggcgtgtgtgagattctta.....tcattgtgtgtaacgtattt 420

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Indexed: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCrus_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	41	9.8	US-08-189-256A-27	Sequence 27, Appl
2	41	9.8	US-09-193-853-27	Sequence 27, Appl
3	41	9.8	US-09-142-114B-7	Sequence 7, Appl
4	39.2	9.3	US-08-189-256A-15	Sequence 15, Appl
5	39.2	9.3	US-09-193-853-15	Sequence 15, Appl
6	36.2	8.6	US-08-189-256A-5	Sequence 5, Appl
7	36.2	8.6	US-09-193-853-5	Sequence 5, Appl
8	34.4	8.2	US-08-189-256A-4	Sequence 4, Appl
9	34.4	8.2	US-09-193-853-4	Sequence 4, Appl
10	34.2	8.1	US-09-336-447A-6	Sequence 6, Appl
11	32.8	7.8	US-08-189-256A-35	Sequence 35, Appl
12	32.8	7.8	US-09-193-853-35	Sequence 35, Appl
13	31.8	7.6	US-08-189-256A-2	Sequence 2, Appl
14	31.8	7.6	US-09-193-853-2	Sequence 2, Appl
15	31.8	7.6	US-09-428-711A-13	Sequence 13, Appl
16	31.6	7.5	US-08-460-93A-5	Sequence 5, Appl
17	31.6	7.5	US-08-782-118-5	Sequence 5, Appl
18	31.6	7.5	US-08-308-872B-3	Sequence 3, Appl
19	31.6	7.5	US-08-308-872B-5	Sequence 5, Appl
20	31.6	7.5	US-08-374-077C-1	Sequence 1, Appl
21	31.6	7.5	US-08-895-590-1	Sequence 1, Appl
22	31.6	7.5	US-09-539-879A-1	Sequence 1, Appl
23	31.2	7.4	US-08-991-789A-251	Sequence 251, App
24	31.2	7.4	US-09-062-451-251	Sequence 251, App
25	30	7.1	US-09-221-017B-473	Sequence 473, App
26	29.8	7.1	US-09-080-983-12	Sequence 12, Appl
27	29.8	7.1	US-09-221-017B-626	Sequence 626, App

28	29.8	7.1	7812	4	US-09-368-590-1	Sequence 1, Appl
29	29.8	7.1	15500	4	US-09-080-983-1	Sequence 1, Appl
30	29.8	7.1	38844	4	US-09-734-675-3	Sequence 3, Appl
31	29.6	7.0	414	4	US-09-042-353-353	Sequence 353, App
32	29.6	7.0	414	4	US-08-758-417A-201	Sequence 201, App
33	29.4	7.0	379	1	US-08-145-617-5	Sequence 5, Appl
34	29.4	7.0	2190	4	US-09-625-188-19	Sequence 19, Appl
35	29.2	7.0	5398	3	US-09-356-952-11	Sequence 11, Appl
36	29	6.9	1727	4	US-09-071-035-295	Sequence 295, App
37	29	6.9	1839	4	US-09-071-035-293	Sequence 293, App
38	28.8	6.9	499	4	US-09-004-838-9	Sequence 9, Appl
39	28.8	6.9	1110	1	US-08-257-341-4	Sequence 4, Appl
40	28.6	6.8	701	3	US-08-331-625A-55	Sequence 55, Appl
41	28.6	6.8	701	4	US-09-494-151-55	Sequence 55, Appl
42	28.6	6.8	777	4	US-08-809-326A-7	Sequence 7, Appl
43	28.6	6.8	813	4	US-08-809-326A-4	Sequence 4, Appl
44	28.6	6.8	1048	4	US-08-809-326A-9	Sequence 9, Appl
45	28.6	6.8	1284	3	US-08-331-625A-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-08-189-256A-27
Sequence 27, Application US/08189256A

Patent No. 5877402

GENERAL INFORMATION:

APPLICANT: Malliga, Pal

APPLICANT: Svab, Zora

APPLICANT: Staub, Jeffrey

APPLICANT: Zouzenko, Oleg V.

APPLICANT: Allison, Lori A.

APPLICANT: Carrier, Helaine

APPLICANT: Kanevski, Ivan

TITLE OF INVENTION: DNA Constructs and Methods for Stably

TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and

TITLE OF INVENTION: Expressing Recombinant Proteins Therein

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman

STREET: 1601 Market Street Suite 720

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/189,256A

FILING DATE: 31-JAN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/111,398

FILING DATE: 25-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/518,763

FILING DATE: 01-MAY-1990

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 1416 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-189-256A-27

Query Match
Best Local Similarity 9.8%; Score 41; DB 2; Length 1416;
Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 28 AGTTGAGGAGGAGGACTTATGTCACCAACAACTAAAGCAAGTGTGATTAA 87
DB 139 AGTTGAGGAGGAGGACTTATGTCACCAACAAAGAGGGATTGAACAGATGATGCAC 198
QY 88 GCTGGTGTAAAG 100
DB 199 GCAGGTTCTCCGG 211

RESULT 2

US-09-193-853-27
Sequence 27, Application US/09193853
Patent No. 6388168

GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Staub, Zora
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carreir, Helaine
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,256
FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-193-853-27

Query Match
Best Local Similarity 9.8%; Score 41; DB 4; Length 1416;
Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 28 AGTTGAGGAGGAGGACTTATGTCACCAACAACTAAAGCAAGTGTGATTAA 87
DB 139 AGTTGAGGAGGAGGACTTATGTCACCAACAAAGAGGGATTGAACAGATGATGCAC 198
QY 88 GCTGGTGTAAAG 100
DB 199 GCAGGTTCTCCGG 211

RESULT 3

US-09-142-114B-7
Sequence 7, Application US/09142114B
Patent No. 6376744

GENERAL INFORMATION:
APPLICANT: Rutgers University
APPLICANT: Maliga, Pal
APPLICANT: Sikdar, Samir R.
APPLICANT: Reddy, Siva Vanga
TITLE OF INVENTION: Plasmid Transformation in Arabidopsis
FILE REFERENCE: 09/142,114
CURRENT APPLICATION NUMBER: US/09/142,114B
CURRENT FILING DATE: 1999-02-05
PRIOR APPLICATION NUMBER: PCT/US97/03444
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1417
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Sequence source:/note="synthetic construct"
Patent No. 6376744
US-09-142-114B-7

Query Match
Best Local Similarity 9.8%; Score 41; DB 4; Length 1417;
Matches 53; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 28 AGTTGAGGAGGAGGACTTATGTCACCAACAACTAAAGCAAGTGTGATTAA 87
DB 140 AGTTGAGGAGGAGGACTTATGTCACCAACAAAGAGGGATTGAACAGATGATGCAC 199
QY 88 GCTGGTGTAAAG 100
DB 200 GCAGGTTCTCCGG 212

RESULT 4

US-08-189-256A-15
Sequence 15, Application US/08189256A
Patent No. 5877402

GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Staub, Zora
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carreir, Helaine
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia

STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,256A
FILING DATE: 31-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/111,398
FILING DATE: 25-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-189-256A-15

Query Match
Best Local Similarity 9.3%; Score 39.2; DB 2; Length 278;
Matches 41; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 234 GTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGTCATG 277

RESULT 5
Sequence 15, Application US/09193853
Patent No. 6388168
GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Staub, Jeffrey
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carter, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
TITLE OF INVENTION: Expressing Recombinant Proteins Therein
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/193,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,256
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-193-853-15

Query Match
Best Local Similarity 9.3%; Score 39.2; DB 4; Length 278;
Matches 41; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 234 GTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGTCATG 277

RESULT 6
US-08-189-256A-5
Sequence 5, Application US/08189256A
Patent No. 5877402
GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Staub, Jeffrey
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carter, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
TITLE OF INVENTION: Expressing Recombinant Proteins Therein
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,256A
FILING DATE: 31-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/111,398
FILING DATE: 25-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-189-256A-5

Query Match
Best Local Similarity 8.6%; Score 36.2; DB 2; Length 191;
Matches . 41; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

28 AGTTTAGGAGGAGGACTTATGTCCACCAACAGAACTAAGCACTG 76
143 AGTTTAGGAGGAGGACTTATGTCCACCAACAGAGGGGAGCGGCTG 191

RESULT 7

US-09-193-853-5
Sequence 5, Application US/09193853
Patent No. 6388168
GENERAL INFORMATION:
APPLICANT: Malliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Staub, Jeffrey
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carrier, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,256
FILING DATE:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-193-853-5

Query Match
Best Local Similarity 8.6%; Score 36.2; DB 4; Length 191;
Matches . 41; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

28 AGTTTAGGAGGAGGACTTATGTCCACCAACAGAACTAAGCACTG 76
Db 143 AGTTTAGGAGGAGGACTTATGTCCACCAACAGAGGGGAGCGGCTG 191

RESULT 8

US-08-189-256A-4
Sequence 4, Application US/08189256A
Patent No. 5877402

GENERAL INFORMATION:
APPLICANT: Malliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Staub, Jeffrey
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carrier, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,256A
FILING DATE: 31-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/111,398
FILING DATE: 25-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 165 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-189-256A-4

Query Match
8.2%; Score 34.4; DB 2; Length 165;

FILED DATE: 25-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-189-256A-35

Query Match
Best Local Similarity 7.8%; Score 32.8; DB 2; Length 52;
Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 29 GTGTAGGAGGAGGACTTATGTACCAACAAGAACTAAGCAAGTGTGG 80
Db 52 GTGTAGGAGGAGGACTTATGTACCAACAAGGAGGAGTGAACAAGATGG 1

RESULT 12
US-09-193-853-35/c
Sequence 35, Application US/09193853
Patent No. 6388168
GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carrier, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
TITLE OF INVENTION: Expressing Recombinant Proteins Therein
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,256
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-193-853-35

Query Match
Best Local Similarity 7.8%; Score 32.8; DB 4; Length 52;
Matches 40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 29 GTGTAGGAGGAGGACTTATGTACCAACAAGAACTAAGCAAGTGTGG 80
Db 52 GTGTAGGAGGAGGACTTATGTACCAACAAGGAGGAGTGAACAAGATGG 1

RESULT 13
US-08-189-256A-2
Sequence 2, Application US/08189256A
Patent No. 5877402

GENERAL INFORMATION:
APPLICANT: Maliga, Pal
APPLICANT: Svab, Zora
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carrier, Helaine
APPLICANT: Kanevski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
TITLE OF INVENTION: Expressing Recombinant Proteins Therein
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,256A
FILING DATE: 31-JAN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/111,398
FILING DATE: 25-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 168 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-189-256A-2

Query Match 7.6%; Score 31.8; DB 2; Length 168;
Best Local Similarity 94.3%; Pred. No. 0.47;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 28 AGTTGTAGGAGGACTTATGTCCACCAACACAGA 62
Db 118 AGTTGTAGGAGGACTTATGTCCACCAACACAGA 152

RESULT 14
US-09-193-853-2
Sequence 2, Application US/09193853
Patent No. 6386168

GENERAL INFORMATION:

APPLICANT: Maligna, Pal
APPLICANT: Svab, Zora
APPLICANT: Staub, Jeffrey
APPLICANT: Zoubenko, Oleg V.
APPLICANT: Allison, Lori A.
APPLICANT: Carer, Helaine
APPLICANT: Kanewski, Ivan
TITLE OF INVENTION: DNA Constructs and Methods for Stably
TITLE OF INVENTION: Transforming Plasmids of Multicellular Plants and
TITLE OF INVENTION: Expressing Recombinant Proteins Therein
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,256
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,763
FILING DATE: 01-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 168 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-193-853-2

Query Match 7.6%; Score 31.8; DB 4; Length 168;
Best Local Similarity 94.3%; Pred. No. 0.47;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 28 AGTTGTAGGAGGACTTATGTCCACCAACACAGA 62
|||||

Db 118 AGTTGTAGGAGGACTTATGTCCACCAACACAGA 152

RESULT 15
US-09-428-711A-13
Sequence 13, Application US/09428711A
Patent No. 6358720

GENERAL INFORMATION:

APPLICANT: Muramatsu, Masaki
APPLICANT: Shirasawa, Takuji
APPLICANT: Tokumitsu, Hiroshi
APPLICANT: No. 6358720uchi, Teruhisa
TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE
FILE REFERENCE: 06501-045001
CURRENT APPLICATION NUMBER: US/09/428,711A
CURRENT FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: PCT/JP98/01246
PRIOR FILING DATE: 1998-03-23
PRIOR APPLICATION NUMBER: JP 9/124798
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 3995
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (388)...(3540)
NAME/KEY: misc_feature
LOCATION: (1)...(3995)
OTHER INFORMATION: n = A,T,C or G
US-09-428-711A-13

Query Match 7.6%; Score 31.8; DB 4; Length 3995;
Best Local Similarity 51.8%; Pred. No. 2;
Matches 72; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 119 ACAACCCGAGTACGAACCAAGATATCTTGGCAGCATTCGAGTAATCTTC 178
Db 2537 ACAGCTACAGAGGAAACCTATGTAGATTCCTCCCTGTGGAGGAGCTGCG 2596
QY 179 AGCTGGGGTCCGCCCTGAAGAAAGCAGAGCTCAGTACGTGGGAATCTTACTGGA 238
Db 2597 ATCCAGAGCTCGGTGGAGGGGCCAGCCAGCCAGACCTGTGATTTACTGTAGGCT 2656
QY 239 CATGACACGTTTGGAC 257
Db 2657 CCCACCCAGGTGGCCAC 2675

Search completed: April 15, 2003, 18:27:33
Job time : 97 secs

QY	6	TTGTGAGATTCCTTAATTCATGAGCTGTAGGGAGGACTTATGTCACCAACGAAAC	65
Db	1239	TTGCACGAAATTTAAATTCATGAGCTGTAGGGAGGAGCTATATCCCAACGAGAGC	1180
QY	66	TAAAGCAAGCTGTGGATTTAAAGCTGCTGTAAAGATTAAATTAACATAACCC	125
Db	1179	TAAAGCAAGCTGTGGATTTAAAGCTGCTGTAAAGATTAAATTAACATAACCTACATAC	1120

QY 126 GGAGTACGAACCAAGATGATGATATCTGGCAGCATTCGATTAACCTCAGCTCG 185
 DB 1119 TGAATATGATCAACCAAGATGATGATATCTGGCAGCATTCGATTAACCTCAGCTCG 1060
 QY 186 GGTTCGCGCTGAAGAAGAGAGAGCGGAGTGTCTGCGAATCTTCTACTGTAATGAGAC 245
 DB 1059 AGTTCACCTGTAAGAAGAGAGAGCGGAGTGTCTGCGAATCTTCTACTGTAATGAGAC 1000
 QY 246 AACTGTGGAGCTGATGAGCTTACCAAGTCTGATGATTAACAAAGAGAGATGATGAT 305
 DB 999 AACTGTGGAGCTGATGAGCTTACCAAGTCTGATGATTAACAAAGAGAGATGATGAT 940
 QY 306 CGAGCCGCTTCCTGGGACCCAGATCAATATCTGTTATGATGATTAACATTAACCT 365
 DB 939 CGAGCCGCTTCCTGGGACCCAGATCAATATCTGTTATGATGATTAACATTAACCT 880
 QY 366 ATTTGAAGAGGCTTCTGATCAATGATGATTAACATTAACCTGATTAACCTGATTA 420
 DB 879 TTTGAAGAGGCTTCTGATCAATGATGATTAACATTAACCTGATTAACCTGATTA 825

RESULT 2

Sequence 3601, Application US/09878574
 Patent No. US20020110548A1
 GENERAL INFORMATION:
 APPLICANT: Byrum, Joseph R.
 APPLICANT: La Rosa, Thomas J.
 APPLICANT: Thompson, Michael D.
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 FILE REFERENCE: 38-21(15401)B
 CURRENT APPLICATION NUMBER: US/09/878,574
 CURRENT FILING DATE: 2001-12-21
 PRIOR APPLICATION NUMBER: 09/333,535
 PRIOR FILING DATE: 1999-06-14
 NUMBER OF SEQ ID NOS: 15775
 SEQ ID NO 3601
 LENGTH: 406
 TYPE: DNA
 ORGANISM: Glycine max
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(406)
 OTHER INFORMATION: unsure at all n locations
 OTHER INFORMATION: Clone ID: LIB3028-008-Q1-B1-C9
 US-09-878-574-3601

Query Match Best Local Similarity 59.7%; Score 250.8; DB 10; Length 406;
 Matches 267; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

DB 5 GTTGTGAGAAATCTTAATTCATGATGATGAGGAGGAGCTTATGTCACCAACAAAGAAA 64
 DB 111 GTTGTGAGAAATCTTAATTCATGATGATGAGGAGGAGCTTATGTCACCAACAAAGAAA 170
 QY 65 CTAAGCAAGTGTGATTAAGCTGTGTTAAGATTAATTAATGATTAATTAATTAATTAAT 124
 DB 171 CTAAGCAAGTGTGATTAAGCTGTGTTAAGATTAATTAATGATTAATTAATTAATTAAT 230
 QY 125 CGAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
 DB 231 CTGACTATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 290
 QY 185 GGGTTCGCGCTGAAGAAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 244
 DB 291 GAGTTCGCGCTGAAGAAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 350
 QY 245 CAACCTGTTGAGTGAAGAGCTTACCAAGTCTTATGATGATTAACAAAGAGAGAGATGATGAT 299
 DB 351 CAACCTGTTGAGTGAAGAGCTTACCAAGTCTTATGATGATTAACAAAGAGAGAGATGATGAT 405

RESULT 3

US-09-070-927A-104
 Sequence 104, Application US/09070927A
 Patent No. US20020120116A1
 GENERAL INFORMATION:
 APPLICANT: Charles A. Kunsch
 APPLICANT: Patrick J. Dillon
 APPLICANT: Steven Barash
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 982
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/070,927A
 FILING DATE: 04-May-2000
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/046,655
 FILING DATE: 1997-05-16
 APPLICATION NUMBER: 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 104:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6433 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 104:
 US-09-070-927A-104

Query Match Best Local Similarity 9.0%; Score 38; DB 10; Length 6433;
 Matches 128; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 3 GCGGTGTGAGAAATCTTAATTCATGATGATGAGGAGGAGCTTATGTCACCAACAAAGAAA 62
 DB 3674 GCGGTGTGAGAAATCTTAATTCATGATGATGAGGAGGAGCTTATGTCACCAACAAAGAAA 3733
 QY 63 AACTAAGCAAGTGTGATTAAGCTGTGTTAAGATTAATTAATGATTAATTAATTAATTAAT 122
 DB 3734 CTGTTAAGCAATGACACCTTACAGGTTGATGATGATGATGATGATGATGATGATGATGAT 3793
 QY 123 CCGGAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 182
 DB 3794 CCGGAGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3853
 QY 183 CGGGTTCGCGCTGAAGAAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 242
 DB 3854 AGTATGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3913
 QY 243 GACAACCTGTTGAGTGAAGAGCTTACCAAGTCTTATGATGATTAACAAAGAGAGAGATGATGAT 280
 DB 3914 GCCTTGTGACCAAGAGCTTTCCTTCAATATGATGATGATGATGATGATGATGATGATGATGAT 3951

Query Match	7.88; Score 32.8; DB 10; Length 495,
-------------	--------------------------------------

```

; SEQ ID NO 1030
; LENGTH: 2373

```

```

; APPLICANT: Reed, Steven G
; APPLICANT: Smith, John M.

```


GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 11:57:09 ; Search time 25043 Seconds
(without alignments)
271.617 Million cell updates/sec

Title: US-09-696-664A-3
Perfect score: 420
Sequence: 1 gcgcgttgagaaattctta.....tcattgtggtaacgatttc 420

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Database: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412.2	98.1	465	12	BG159568 OV2_6_B06
2	412.2	98.1	488	10	AM285810 LG1_237_A
3	412.2	98.1	490	10	BE598732 P11_81_A0
4	412.2	98.1	490	12	BG053104 RH12_16_
5	412.2	98.1	492	12	BG559375 RH12_54_
6	412.2	98.1	495	10	AM563595 LG1_241_C

7	412.2	98.1	496	12	BG102147 RH12_21_
8	412.2	98.1	517	13	BM329396 P1C1_30_B
9	412.2	98.1	536	12	BG465717 RH12_47_
10	412.2	98.1	546	10	BE596352 P11_49_C0
11	412.2	98.1	551	10	BE596365 P11_52_A0
12	412.2	98.1	552	12	BF705205 RH12_2_A
13	412.2	98.1	555	12	BM323693 P1C1_22_C
14	412.2	98.1	576	10	AM671847 LG1_352_C
15	412.2	98.1	576	12	BG102878 RH12_34_
16	412.2	98.1	582	12	BF657539 OV2_23_E0
17	412.2	98.1	593	13	BM326083 P1C1_65_E
18	412.2	98.1	598	13	BG944809 LG1_266_E
19	412.2	98.1	599	13	BM326559 P1C1_58_E
20	412.2	98.1	601	13	BM326409 P1C1_56_F
21	412.2	98.1	614	10	BE594480 P11_33_H0
22	412.2	98.1	615	10	BE599555 P1C1_41_D
23	412.2	98.1	630	12	BF704811 OV2_31_C0
24	412.2	98.1	630	13	BM325003 P1C1_38_E
25	411	97.9	515	13	BM325550 P1C1_46_A
26	411	97.9	627	13	BM325185 P1C1_41_D
27	410.6	97.8	583	10	BE592098 LG1_224_C
28	410.4	97.7	596	12	BG357871 OV2_31_C0
29	408.2	97.2	444	12	BG560424 RH12_74_
30	408.2	97.2	450	12	BG241641 RH12_49_
31	408.2	97.2	487	12	BG465536 RH12_45_
32	408.2	97.2	487	13	BM323586 P1C1_21_A
33	408.2	97.2	488	10	AM671318 LG1_335_B
34	408.2	97.2	490	12	BG159586 OV2_6_D09
35	408.2	97.2	491	10	AM564022 LG1_281_G
36	408.2	97.2	491	12	BG465556 RH12_46_
37	408.2	97.2	513	12	BG559241 RH12_52_
38	408.2	97.2	513	13	BM322939 P1C1_15_D
39	408.2	97.2	514	13	BM323531 P1C1_20_D
40	408.2	97.2	562	12	BG052307 RH12_12_
41	408.2	97.2	562	13	BM324614 P1C1_33_D
42	408.2	97.2	572	12	BG488245 RH12_61_
43	406.6	96.8	487	10	AM671457 LG1_347_B
44	406.6	96.8	487	10	BE597559 P11_71_B0
45	406.6	96.8	488	10	AM286737 LG1_204_G

ALIGNMENTS

RESULT 1
LOCUS BG159568 465 bp mRNA linear EST 06-FEB-2001
DEFINITION OV2_6_B06.b1.A002 Ovary 2 (OV2) Sorghum bicolor CDNA, mRNA
ACCESSION BG159568
VERSION BG159568.1 GI:12693232
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 465)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: ovaries of varying immature stages
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 435
POLA-No.
Location/Qualifiers

FEATURES

source

1. 465
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 2 (OV2)"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 124 a 94 c 113 g 134 t
ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 465;
Best Local Similarity 99.3%; Pred. No. 4.7e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGTTGAGAAATCTTAATTCATGAGTGTGAGGAGGAGCTTATGTCACCAACAGAA 63
DB 11 CGTTGAGAAATCTTAATTCATGAGTGTGAGGAGGAGCTTATGTCACCAACAGAA 70

QY 64 ACTAAGCAAGTGTGATTTAAAGCTGGTGTAAAGATTAAATGACTTACTACACC 123
DB 71 ACTAAGCAAGTGTGATTTAAAGCTGGTGTAAAGATTAAATGACTTACTACACC 130

QY 124 CCGGAGTACGAACCAAGATACATATCTTGGCAGCATTCGAGTACTCTCAGCTC 183
DB 131 CCGGAGTACGAACCAAGATACATATCTTGGCAGCATTCGAGTACTCTCAGCTC 190

QY 184 GGGGTCCGCTGAAGAAGCAGAGCTGACATGCTGCGAGATCTTACTGTACATGG 243
DB 191 GGGGTCCGCTGAAGAAGCAGAGCTGACATGCTGCGAGATCTTACTGTACATGG 250

QY 244 ACACTGTTGGACTGATGACTTACCAAGCTTGTATCGTTACAAAGAGATGCTATCAC 303
DB 251 ACACTGTTGGACTGATGACTTACCAAGCTTGTATCGTTACAAAGAGATGCTATCAC 310

QY 304 ATCGAGCCGCTCCGGGAGCCAGATCAATATATCTGTATGATGCTTATGATGCTTATGAC 363
DB 311 ATCGAGCCGCTCCGGGAGCCAGATCAATATATCTGTATGATGCTTATGATGCTTATGAC 370

QY 364 CTATTGGAAGAGGGTCTTGTACTACATGTTTACTTCTTATGTTGGGTAACGTAATT 420
DB 371 CTATTGGAAGAGGGTCTTGTACTACATGTTTACTTCTTATGTTGGGTAACGTAATT 427

RESULT 2 488 bp mRNA linear EST 19-JUL-2000
AM285810 LGI_237_A11_b1_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
LOCUS
DEFINITION
SEQUENCE
ACCESSION
AM285810
VERSION
AM285810.2 GI:6858276
KEYWORDS
EST.
SOURCE
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 488)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L. H.
An EST database from Sorghum: light-grown seedlings
Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6675654.
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmp@pratt.uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV
High quality sequence stop: 487
POLYA-No.

FEATURES
source
location/Qualifiers
1. 488
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LGI)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 131 a 101 c 112 g 144 t
ORIGIN

Query Match 98.1%; Score 412.2; DB 10; Length 488;
Best Local Similarity 99.3%; Pred. No. 4.8e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGTTGAGAAATCTTAATTCATGAGTGTGAGGAGGAGCTTATGTCACCAACAGAA 63
DB 3 CGTTGAGAAATCTTAATTCATGAGTGTGAGGAGGAGCTTATGTCACCAACAGAA 62

QY 64 ACTAAGCAAGTGTGATTTAAAGCTGGTGTAAAGATTAAATGACTTACTACACC 123
DB 63 ACTAAGCAAGTGTGATTTAAAGCTGGTGTAAAGATTAAATGACTTACTACACC 122

QY 124 CCGGAGTACGAACCAAGATACATATCTTGGCAGCATTCGAGTACTCTCAGCTC 183
DB 123 CCGGAGTACGAACCAAGATACATATCTTGGCAGCATTCGAGTACTCTCAGCTC 182

QY 184 GGGGTCCGCTGAAGAAGCAGAGCTGACATGCTGCGAGATCTTACTGTACATGG 243
DB 183 GGGGTCCGCTGAAGAAGCAGAGCTGACATGCTGCGAGATCTTACTGTACATGG 242

QY 244 ACACTGTTGGACTGATGACTTACCAAGCTTGTATCGTTACAAAGAGATGCTATCAC 303
DB 243 ACACTGTTGGACTGATGACTTACCAAGCTTGTATCGTTACAAAGAGATGCTATCAC 302

QY 304 ATCGAGCCGCTCCGGGAGCCAGATCAATATATCTGTATGATGCTTATGATGCTTATGAC 363
DB 303 ATCGAGCCGCTCCGGGAGCCAGATCAATATATCTGTATGATGCTTATGATGCTTATGAC 362

QY 364 CTATTGGAAGAGGGTCTTGTACTACATGTTTACTTCTTATGTTGGGTAACGTAATT 420
DB 363 CTATTGGAAGAGGGTCTTGTACTACATGTTTACTTCTTATGTTGGGTAACGTAATT 419

RESULT 3 490 bp mRNA linear EST 18-AUG-2000
BE598732 P11_B1_A06_b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
LOCUS
DEFINITION
SEQUENCE
ACCESSION
BE598732
VERSION
BE598732.1 GI:9853805
KEYWORDS
EST.
SOURCE
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 490)
Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt, L. H.
An EST database from Sorghum: pathogen-induced plants
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210

Email: mmp@pratt.edu
 Sequences have been trimmed to exclude PolyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 465
 POLYA=NO.

FEATURES
 source Location/Qualifiers

1. 490
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (PI1)"
 /note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Two-week-old sorghum plants (BX 623 cultivar) were infected with pathogen (isolate PM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
 BASE COUNT 132 a 100 c 116 g 141 t 1 others
 ORIGIN

Query Match 98.1%; Score 412.2; DB 10; Length 490;
 Best Local Similarity 99.3%; Pred. No. 4.8e-118;
 Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGTTGTGAGAAATCTTAATCATAGTGTGAGAGGAGCTTATGTCACCAACAGAA 63
 DB 11 CGTTGTGAGAAATCTTAATCATAGTGTGAGAGGAGCTTATGTCACCAACAGAA 70
 QY 64 ACTAAGCAAGTGTGATTAAGCTGTGATTAAGATTAATTAATGACTTACTACACC 123
 DB 71 ACTAAGCAAGTGTGATTAAGCTGTGATTAAGATTAATTAATGACTTACTACACC 130
 QY 124 CCGAGTAGAAGCAAGCAAGATGATATCTTGGCAGCATTCGAGTAATCTCTAGCTC 183
 DB 131 CCGAGTAGAAGCAAGCAAGATGATATCTTGGCAGCATTCGAGTAATCTCTAGCTC 190
 QY 184 GGGGTCCGCTGAGAAAGCAGAGCTGAGTAATCTTCTAGTGTACATGG 243
 DB 191 GGGGTCCGCTGAGAAAGCAGAGCTGAGTAATCTTCTAGTGTACATGG 250
 QY 244 ACAACTGTTGGAGTGAATGACATGCTGATCTGTTCAAAAGAGCATGATAC 303
 DB 251 ACAACTGTTGGAGTGAATGACATGCTGATCTGTTCAAAAGAGCATGATAC 310
 QY 304 ATCGAGCCGCTCTCTGGGAGCCAGATCAATATCTGTATGATGCTTATCATTAAC 363
 DB 311 ATCGAGCCGCTCTCTGGGAGCCAGATCAATATCTGTATGATGCTTATCATTAAC 370
 QY 364 CTATTGGAAGAGGTTCTGTACTAATCATGTTACTTCAATGCGGTAAGTATT 420
 DB 371 CTATTGGAAGAGGTTCTGTACTAATCATGTTACTTCAATGCGGTAAGTATT 427

RESULT 4
 BG053104 490 bp mRNA linear EST 25-JAN-2001
 LOCUS RH1Z2_16.F07.b1_A003 Rhizome2 (RH1Z2) Sorghum propinquum cDNA, mRNA
 DEFINITION
 sequence.
 ACCESSION BG053104
 VERSION BG053104.1 GI:12508448
 KEYWORDS EST.

SOURCE
 ORGANISM Sorghum propinquum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 490)
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudan,M. and Pratt,L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt.edu
 Seq primer: JEN REV
 High quality sequence stop: 463
 POLYA=NO.

FEATURES
 source Location/Qualifiers

1. 490
 /organism="Sorghum propinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH1Z2)"
 /note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 132 a 101 c 116 g 141 t
 ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 490;
 Best Local Similarity 99.3%; Pred. No. 4.8e-118;
 Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGTTGTGAGAAATCTTAATCATAGTGTGAGAGGAGCTTATGTCACCAACAGAA 63
 DB 11 CGTTGTGAGAAATCTTAATCATAGTGTGAGAGGAGCTTATGTCACCAACAGAA 70
 QY 64 ACTAAGCAAGTGTGATTAAGCTGTGATTAAGATTAATTAATGACTTACTACACC 123
 DB 71 ACTAAGCAAGTGTGATTAAGCTGTGATTAAGATTAATTAATGACTTACTACACC 130
 QY 124 CCGAGTAGAAGCAAGCAAGATGATATCTTGGCAGCATTCGAGTAATCTCTAGCTC 183
 DB 131 CCGAGTAGAAGCAAGCAAGATGATATCTTGGCAGCATTCGAGTAATCTCTAGCTC 190
 QY 184 GGGGTCCGCTGAGAAAGCAGAGCTGAGTAATCTTCTAGTGTACATGG 243
 DB 191 GGGGTCCGCTGAGAAAGCAGAGCTGAGTAATCTTCTAGTGTACATGG 250
 QY 244 ACAACTGTTGGAGTGAATGACATGCTGATCTGTTCAAAAGAGCATGATAC 303
 DB 251 ACAACTGTTGGAGTGAATGACATGCTGATCTGTTCAAAAGAGCATGATAC 310
 QY 304 ATCGAGCCGCTCTCTGGGAGCCAGATCAATATCTGTATGATGCTTATCATTAAC 363
 DB 311 ATCGAGCCGCTCTCTGGGAGCCAGATCAATATCTGTATGATGCTTATCATTAAC 370
 QY 364 CTATTGGAAGAGGTTCTGTACTAATCATGTTACTTCAATGCGGTAAGTATT 420
 DB 371 CTATTGGAAGAGGTTCTGTACTAATCATGTTACTTCAATGCGGTAAGTATT 427

RESULT 5
 BG559375 492 bp mRNA linear EST 10-APR-2001
 LOCUS RH1Z2_54.A11.b1_A003 Rhizome2 (RH1Z2) Sorghum propinquum cDNA, mRNA
 DEFINITION
 sequence.
 ACCESSION BG559375
 VERSION BG559375.1 GI:13588373
 KEYWORDS EST.

	RESULT	6
AWS63595		
LOCUS		
DEFINITION		
AWS63595 495 bp mRNA linear EST 19-JUL-2000		
IGT_241_c12_b1_A002 light grown 1 (IGI) Sorghum bicolor CDNA, mRNA		
sequence.		
AWS63595		
ACCESSION		
VERSION		
AWS63595.1 GI:721473		
KEYWORDS		
EST.		

RESULT 7	496 bp	mRNA	linear	EST 30-JAN-2001
BG102147				
LOCUS				
DEFINITION				
	BG102147			
	RHIZ2.21.D07.b1.A003	Rhizome2 (RHIZ2)	Sorghum prolinguam	cdna, mRNA
				Sequence.

ACCESSION	BGI02147
VERSION	BGI02147.1
KEYWORDS	EST.
SOURCE	Sorghum propinquum.
ORGANISM	Sorghum propinquum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 496) Cordonier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt, L.H.
TITLE	An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 543 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Seq primer: JEN REV High quality sequence stop: 488 POLYA-No.
FEATURES	location/Qualifiers
source	1..496 /organism="Sorghum propinquum" /db_xref="taxon:132711" /clone_1lb="Rhizome2 (RHIZ2)" /note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from polyA RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	133 a 103 c 116 g 144 t
ORIGIN	
Query Match	98.1%; Score 412.2; DB 12; Length 496;
Best Local Similarity	99.3%; Pred. No. 4.8e-118;
Matches 414:	Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY	4 CGTGTGAGAAATCTTAATTCATGATTGTAGAGAGCGACTATGTCCACAACAGCA 63
DB	11 CGTTTGAGAAATCTTAATTCATGATTGTAGAGAGCGACTATGTCCACAACAGCA 70
OY	64 ACTAAGCAAGTGTGGATTAAAGCTGGTTAAGATTATAATTTGACTTACTACACC 123
DB	71 ACTAAGCAAGTGTGGATTAAAGCTGGTTAAGATTATAATTTGACTTACTACACC 130
OY	124 CCGGAGTAGGAACCAAGATCTGATATCTTTGGAGCATTTCCGAGTAACTCTCACTC 183
DB	131 CCGGAGTAGGAACCAAGATCTGATATCTTTGGAGCATTTCCGAGTAACTCTCACTC 190
OY	184 GGGGTTCGCCGAGAAGACAGAGCTGCAGTAGCTGCGGANTCTTCTACTGTCATG 243
DB	191 GGGGTTCGCCGAGAAGACAGAGCTGCAGTAGCTGCGGANTCTTCTACTGTCATG 250
OY	244 ACAACTGTTGGAGTGTGACTTACCACTCTTGATCGTTACAAAAGAGATGCTATCAC 303
DB	251 ACAACTGTTGGAGTGTGACTTACCACTCTTGATCGTTACAAAAGAGATGCTATCAC 310
OY	304 ATCGAGCCCCGTTCCTGGGAGCCAGATCAATATCTGTATGTAGTACTTATCCATTAGAC 363
DB	311 ATCGAGCCCCGTTCCTGGGAGCCAGATCAATATCTGTATGTAGTACTTATCCATTAGAC 370
OY	364 CTATTGTGAAGAGGGTCTGTACTTAACATGTTTACTTTCATGTTGGGTAAAGCTATT 420
DB	371 CTATTGTGAAGAGGGTCTGTACTTAACATGTTTACTTTCATGTTGGGTAAAGCTATT 427
RESULT 8	
LOCUS	BM323998
DEFINITION	BM323998 517 bp mRNA linear EST 04-JAN-2002 P101_30_B01.b1_A002 Pathogen-infected compatible 1 (P1C1) Sorghum bicolor cDNA, mRNA sequence.

ACCESSION	BM323998
VERSION	BM323998.1
KEYWORDS	GI:18062304
SOURCE	EST.
ORGANISM	sorghum, Sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
AUTHORS	1 (bases 1 to 517) Cordonnier-Pratt,M.-M., Gingle,A., Fang,G.C., Dean,R., Wing,R., Sudman,M. and Pratt,L.H.
TITLE	An EST database from Sorghum: plants infected with a compatible pathogen
JOURNAL	Unpublished (2002)
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector, and regions below Phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with polyTmix or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV High quality sequence stop: 491 POLYA-No.
FEATURES	location/Qualifiers 1..517 /organism="Sorghum bicolor" /cultivar="BTx623" /db_xref="taxon:4558" /clone_lib="Pathogen-infected compatible 1 (PIC1)" /tissue_type="leaves" /der_stage="4-week-old seedlings infected with Colletotrichum graminicola" /note="Vector: Bluescript II SK(-) from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old RM421, a sorghum isolate of the anthracnose pathogen Colletotrichum graminicola. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT	137 a 111 c 121 g 148 t
ORIGIN	
Query Match	98.1%; Score 412.2; DB 13; Length 517;
Best Local Similarity	99.3%; Pred. NO. 4.9e-118;
Matches 414; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
4	CGTTGGAGATTTCTTAATCATGAGTGTAGGAGGAGCTTATGTCACCACAAACAGAA 63
11	CGTTGTGAAATTCCTTAATCATGAGTGTAGGAGGAGGAGCTTATGTCACCAACAAACAGAA 70
64	ACTAAGCAAGGTGGATTTAAAGCTGTGTTAAGAGATTATAAAATGCTACTACAC 123
71	ACTAAGCAAGGTGGATTTAAAGCTGTGTTAAGAGATTATAAAATGCTACTACAC 130
124	CCGAGTAGCAACCAAGATACTGATACCTTGGCAGCATTCGAGTACTCTCAGCTC 183
131	CGGAGTAGCAACCAAGATACTGATACCTTGGCAGCATTCGAGTACTCTCAGCTC 190
184	GGGGTTCGGCTGGAAGACAGAGAGCTGAGTAGCTCGGAACTTTCTACTGTCATG 243

Db 191 GGGGTTCCGCTGAAGAACAGAGAGCTGACAGTCTGCGGAATCTTCTACTGATACATGG 250

QY 244 ACAACTGTTGGAGTGAATGAGTACCAAGTCTTGAATCTTGAACAAAGAGAGATGATAC 303

Db 251 ACAACTGTTGGAGTGAATGAGTACCAAGTCTTGAATCTTGAACAAAGAGAGATGATAC 310

QY 304 ATCGAGCCGCTTCTGGGAGCCAGATCAATATATCTGTTATGATGATTCATTCATAC 363

Db 311 ATCGAGCCGCTTCTGGGAGCCAGATCAATATATCTGTTATGATGATTCATTCATAC 370

QY 364 CTATTGAAGAGGTTCTGTTACTTAACAGTTTACTTCAATGCGGAATCTTCTACTGATAC 420

Db 371 CTATTGAAGAGGTTCTGTTACTTAACAGTTTACTTCAATGCGGAATCTTCTACTGATAC 427

RESULT 9
Bg465717 536 bp mRNA linear EST 20-MAR-2001
LOCUS Bg465717
DEFINITION RH122_47_G09.b1_A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
ACCESSION Bg465717
VERSION Bg465717.1 GI:13394693
KEYWORDS EST.
SOURCE Sorghum prolinguam.
ORGANISM Sorghum prolinguam.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 536)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt, L. H.
TITLE An EST database from Sorghum: Sorghum prolinguam rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 526
POLYA-No.

FEATURES
source location/Qualifiers
1..536
/organism="Sorghum prolinguam"
/db_xref="taxon:132711"
/clone_11b="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: Bluescript II from Lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
USE COUNT 145 a 112 c 127 g 152 t
ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 536;
Best Local Similarity 99.3%; Pred. No. 4.9e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGCTTATGTCACCAACAGAA 63

Db 11 CGTTGTGAGAAATCTTAATTCATGAGTTGTAGGAGGAGCTTATGTCACCAACAGAA 70

QY 64 ACTAAGCAAGTGTGATTTAAAGCTGGTGAAGATTATAATGACTTACTACAC 123

Db 71 ACTAAGCAAGTGTGATTTAAAGCTGGTGAAGATTATAATGACTTACTACAC 130

QY 124 CCGAGTACGAAACCAAGATGATATCTGGACAGATTCGAGTAATCTCTACACTC 183

Db 131 CCGAGTACGAAACCAAGATGATATCTGGACAGATTCGAGTAATCTCTACACTC 190

QY 184 GGGGTTCCGCTGAAGAACAGAGAGCTGACAGTCTGCGGAATCTTCTACTGATACATGG 243

Db 191 GGGGTTCCGCTGAAGAACAGAGAGCTGACAGTCTGCGGAATCTTCTACTGATACATGG 250

QY 244 ACAACTGTTGGAGTGAATGAGTACCAAGTCTTGAATCTTGAACAAAGAGAGATGATAC 303

Db 251 ACAACTGTTGGAGTGAATGAGTACCAAGTCTTGAATCTTGAACAAAGAGAGATGATAC 310

QY 304 ATCGAGCCGCTTCTGGGAGCCAGATCAATATATCTGTTATGATGATTCATTCATAC 363

Db 311 ATCGAGCCGCTTCTGGGAGCCAGATCAATATATCTGTTATGATGATTCATTCATAC 370

QY 364 CTATTGAAGAGGTTCTGTTACTTAACAGTTTACTTCAATGCGGAATCTTCTACTGATAC 420

Db 371 CTATTGAAGAGGTTCTGTTACTTAACAGTTTACTTCAATGCGGAATCTTCTACTGATAC 427

RESULT 10
BE596352 546 bp mRNA linear EST 18-AUG-2000
LOCUS BE596352
DEFINITION P11_49_G09.b1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE596352
VERSION BE596352.1 GI:9851425
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 546)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt, L. H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 494
POLYA-No.

FEATURES
source location/Qualifiers
1..546
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_11b="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: Bluescript II from Lambda Zap II; Site: 1: XhoI; Site: 2: EcoRI; Two-week-old sorghum plants (B7x 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
USE COUNT 151 a 113 c 128 g 154 t
ORIGIN

Query Match 98.1%; Score 412.2; DB 10; Length 546;
Best Local Similarity 99.3%; Pred. No. 5e-118;

High quality sequence stop: 502
POLYA-NO.

FEATURES

source

Location/Qualifiers

1..552
/organism="Sorghum propinquum"
/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: Bluescript II from lambda
zap II; Site: 1: XhoI; Site: 2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT

151 a 115 c 131 g 155 t

ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 552;
Best Local Similarity 99.3%; Pred. No. 5e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

4 CGTTGAGAAATTTCTAATTCATGATGTTGAGGAGGAGCTTAATGTCACCAACAGAA 63
11 CGTTGAGAAATTTCTAATTCATGATGTTGAGGAGGAGCTTAATGTCACCAACAGAA 70

64 ACTAAGCAAGTGTGATTAAGCTGTAAAGATTAATTAATGACTTACTACACC 123
71 ACTAAGCAAGTGTGATTAAGCTGTAAAGATTAATTAATGACTTACTACACC 130

124 CCGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAATCTCAGCTC 183
131 CCGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAATCTCAGCTC 190

184 GGGGTCCGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
191 GGGGTCCGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 250

244 ACAACTGTTGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 303
251 ACAACTGTTGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310

304 ATGAGGCCCTCTCTGGGAGCCGAGATCATATATCTGTATGATGATGATGATGATGAT 363
311 ATGAGGCCCTCTCTGGGAGCCGAGATCATATATCTGTATGATGATGATGATGATGAT 370

364 CTATTGAGAGGGTCTTGTACTAATCATGTTTCAATGAGGATGATGATGATGATGAT 420
371 CTATTGAGAGGGTCTTGTACTAATCATGTTTCAATGAGGATGATGATGATGATGAT 427

RESULT 13
BM323693 565 bp mRNA linear EST 04-JAN-2002
LOCUS PIC1.22.C01.b1.A002 Pathogen-Infected compatible 1 (PIC1) Sorghum
bicolor cDNA, mRNA sequence.

DEFINITION BM323693
VERSION BM323693.1 GI:18061719
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 565)
Cordonier-Pratt, M.-M., Gingle, A., Fang, G.C., Dean, R., Wing, R.,
Sudman, M. and Pratt, L.H.
An EST database from Sorghum: plants infected with a compatible
pathogen
Unpublished (2002)
Contact: Cordonier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Km. 2502, Athens, GA 30602-1271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

JOURNAL
COMMENT

REFERENCE
AUTHORS

TITLE

Sequences have been trimmed to exclude Polya, vector, and regions
below phred quality 16. The threshold for highest quality sequence
is 20. Three prime sequences, which are obtained with PolyTmix or
T7 sequencing primer, are presented as the reverse complement.
Seq primer: JEN REV
High quality sequence stop: 500
POLYA-NO.

FEATURES

source

Location/Qualifiers

1..565
/organism="Sorghum bicolor"
/db_xref="taxon:4558"

/clone_lib="Pathogen-Infected compatible 1 (PIC1)"
/dev_stage="4-week-old seedlings infected with
Colletotrichum graminicola"
/note="Vector: Bluescript II SK(-) from lambda zap II;
Site: 1: XhoI; Site: 2: EcoRI; Four-week-old sorghum
seedlings were sprayed with spore suspension prepared from
3-week-old FRM421, a sorghum isolate of the anthracnose
pathogen Colletotrichum graminicola. Inoculated plants
were kept in a 25 C dark growth chamber with 100% relative
humidity for 24 hr, followed by 12/12 hr of light/dark
cycle at 25 C with 90% relative humidity for another 24
hr. All leaves were harvested and quick frozen with liquid
nitrogen and stored in a -80 C freezer. The library was
made from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision.
WARNING: While most or all ESTs are expected to derive
from the host plant, no effort was made to eliminate ESTs
deriving from the pathogen."

BASE COUNT 152 a 117 c 135 g 161 t

Query Match 98.1%; Score 412.2; DB 13; Length 565;
Best Local Similarity 99.3%; Pred. No. 5e-118;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

4 CGTTGAGAAATTTCTAATTCATGATGTTGAGGAGGAGCTTAATGTCACCAACAGAA 63
11 CGTTGAGAAATTTCTAATTCATGATGTTGAGGAGGAGCTTAATGTCACCAACAGAA 70

64 ACTAAGCAAGTGTGATTAAGCTGTAAAGATTAATTAATGACTTACTACACC 123
71 ACTAAGCAAGTGTGATTAAGCTGTAAAGATTAATTAATGACTTACTACACC 130

124 CCGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAATCTCAGCTC 183
131 CCGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAATCTCAGCTC 190

184 GGGGTCCGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
191 GGGGTCCGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 250

244 ACAACTGTTGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 303
251 ACAACTGTTGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310

304 ATGAGGCCCTCTCTGGGAGCCGAGATCATATATCTGTATGATGATGATGATGATGAT 363
311 ATGAGGCCCTCTCTGGGAGCCGAGATCATATATCTGTATGATGATGATGATGATGAT 370

364 CTATTGAGAGGGTCTTGTACTAATCATGTTTCAATGAGGATGATGATGATGATGAT 420
371 CTATTGAGAGGGTCTTGTACTAATCATGTTTCAATGAGGATGATGATGATGATGAT 427

RESULT 14
AM671847 576 bp mRNA linear EST 19-JUL-2000
LOCUS LGL352.C10.b1.A002 Light grown 1 (LGL) Sorghum bicolor cDNA, mRNA
sequence.

DEFINITION AM671847
ACCESSION AM671847

COMMENT

REFERENCE

AUTHORS

TITLE

VERSION AM671847.1 GI:7535747
 EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC
 clade: Panicoideae: Andropogoneae: Sorghum.
 1 (bases 1 to 576)
 Cordonier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.
 An EST database from Sorghum: light-grown seedlings
 Unpublished (2000)
 CONTACT: Cordonier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt.uga.edu
 Sequences have been trimmed to exclude polyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 490
 POLYA-No.

FEATURES

source location/Qualifiers
 1..576
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Light Grown 1 (LG1)"
 /note="Organ: 10- to 14-day-old light-grown (greenhouse)
 seedlings; Vector: lambda Zap; Site.1: XhoI; Site.2: EcoRI
 ; The library was made from poly-A RNA in the cloning
 vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 156 a 112 c 122 g 186 t
 ORIGIN

Query Match 98.1%; Score 412.2; DB 10; Length 576;
 Best Local Similarity 99.3%; Pred. No. 5.1e-118;
 Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Db 4 CGTTGAGAAATCTTAATCATGAGTGTGAGGAGGACTTATGTCACCAACAGAA 63
 131 CGTTGAGAAATCTTAATCATGAGTGTGAGGAGGACTTATGTCACCAACAGAA 190
 64 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGGATTATTAATGACTTACTACACC 123
 191 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGGATTATTAATGACTTACTACACC 250
 QY 124 CCGGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAACCTCTCAGCTC 183
 Db 251 CCGGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAACCTCTCAGCTC 310
 QY 184 GGGGTTCCGCTGAGAGAGAGAGCTGAGTGCAGGAACTCTTACTGCTACATGG 243
 Db 311 GGGGTTCCGCTGAGAGAGAGAGCTGAGTGCAGGAACTCTTACTGCTACATGG 370
 QY 244 ACAAGTGTGAGCTGATGAGTACCAAGTCTGATCGTTACAAAGAGAGATGCTAC 303
 Db 371 ACAAGTGTGAGCTGATGAGTACCAAGTCTGATCGTTACAAAGAGAGATGCTAC 430
 QY 304 ATCGAGCCCGTCTCTGGGAGCCAGATCAATATATCTGTATGTAGCTTATCCATTAGAC 363
 Db 431 ATCGAGCCCGTCTCTGGGAGCCAGATCAATATATCTGTATGTAGCTTATCCATTAGAC 490
 QY 364 CTATTGGAAGAGGTTCTGTACTAACAATGTTACTTCAATGTTGGTAACGTAATT 420
 Db 491 CTATTGGAAGAGGTTCTGTACTAACAATGTTACTTCAATGTTGGTAACGTAATT 547

RESULT 15
 BG102878 576 bp mRNA linear EST 30-JAN-2001
 LOCUS BG102878

DEFINITION RHIZ2.34.E06.D1.A003 Rhizome2 (RHIZ2) Sorghum prolinquum cDNA, mRNA
 sequence.
 ACCESSION BG102878
 VERSION BG102878.1 GI:12617711
 EST.
 KEYWORDS
 SOURCE Sorghum prolinquum.
 ORGANISM Sorghum prolinquum
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC
 clade: Panicoideae: Andropogoneae: Sorghum.
 1 (bases 1 to 576)
 Cordonier-Pratt, M.-M., Gingle, A., Peterson, A., Sudan, M. and Pratt
 , L.H.
 An EST database from Sorghum: Sorghum prolinquum rhizomes
 Unpublished (2000)
 CONTACT: Cordonier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt.uga.edu
 Seq primer: JEN REV
 High quality sequence stop: 494
 POLYA-No.

FEATURES

source location/Qualifiers
 1..576
 /organism="Sorghum prolinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RHIZ2)"
 /note="Organ: Rhizomes; Vector: pBluescript II from lambda
 Zap II; Site.1: XhoI; Site.2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 156 a 118 c 136 g 166 t
 ORIGIN

Query Match 98.1%; Score 412.2; DB 12; Length 576;
 Best Local Similarity 99.3%; Pred. No. 5.1e-118;
 Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Db 4 CGTTGAGAAATCTTAATCATGAGTGTGAGGAGGACTTATGTCACCAACAGAA 63
 11 CGTTGAGAAATCTTAATCATGAGTGTGAGGAGGACTTATGTCACCAACAGAA 70
 QY 64 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGGATTATTAATGACTTACTACACC 123
 Db 71 ACTAAGCAAGTGTGATTTAAAGCTGCTTAAAGGATTATTAATGACTTACTACACC 130
 QY 124 CCGGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAACCTCTCAGCTC 183
 Db 131 CCGGAGTACGAACCAAGATGATATCTTGGCAGCATTCGAGTAACCTCTCAGCTC 190
 QY 184 GGGGTTCCGCTGAGAGAGAGAGCTGAGTGCAGGAACTCTTACTGCTACATGG 243
 Db 191 GGGGTTCCGCTGAGAGAGAGAGCTGAGTGCAGGAACTCTTACTGCTACATGG 250
 QY 244 ACAAGTGTGAGCTGATGAGTACCAAGTCTGATCGTTACAAAGAGAGATGCTAC 303
 Db 251 ACAAGTGTGAGCTGATGAGTACCAAGTCTGATCGTTACAAAGAGAGATGCTAC 310
 QY 304 ATCGAGCCCGTCTCTGGGAGCCAGATCAATATATCTGTATGTAGCTTATCCATTAGAC 363
 Db 311 ATCGAGCCCGTCTCTGGGAGCCAGATCAATATATCTGTATGTAGCTTATCCATTAGAC 370
 QY 364 CTATTGGAAGAGGTTCTGTACTAACAATGTTACTTCAATGTTGGTAACGTAATT 420
 Db 371 CTATTGGAAGAGGTTCTGTACTAACAATGTTACTTCAATGTTGGTAACGTAATT 427

Search completed: April 16, 2003, 01:25:10
 Job time : 25049 secs

